Faculdade de Ciências da Universidade do Porto

The evolution of the melanopsin gene family (*OPN4m* and *OPN4x*) in vertebrates

Rui Carlos Pinto Borges

Dissertação de Mestrado em Biodiversidade, Genética e Evolução

2012

Rui Carlos Pinto Borges

The evolution of the melanopsin gene family (*OPN4m* and *OPN4x*) in vertebrates

Dissertação de Candidatura ao grau de Mestre em Biodiversidade, Genética e Evolução submetida à Faculdade de Ciências da Universidade do Porto

Orientador	Prof. Doutor Agostinho Antunes
Categoria	Investigador Auxiliar e Professor Auxiliar Convidado
Afiliação	Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR) e Faculdade de Ciências da Universidade do Porto (FCUP)

Co-orientador	Prof. Doutor Vitor Vasconcelos
Categoria	Professor Catedrático
Afiliação	Faculdade de Ciências da Universidade do Porto (FCUP)

There is grandeur in this view of life […], whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.
Charles Darwin

Natural selection is not evolution.
Ronald Fisher

Acknowledgements

To carry out this work many people and institutions have helped me and to those I want to express my gratitude.

Firstly, I would like to thank my supervisor, Professor Agostinho Antunes, for his supervision and advices, friendship and enthusiasm during the development of this thesis. Also, for the full commitment with my academic goals, as well as the enormous trust placed on me, that was very significant. I would also like to thank my co-advisor, Professor Vitor Vasconcelos for welcoming me into his group with such warmth.

To all LEGE staff I would like to thank the good working environment that has always provided. Specially, I am grateful to João Machado for teaching me the basic theoretical concepts in molecular evolution, for helping me clear up and question my results and also for the discussions about the limits of the state-of-the-art theories. To Kartik Sunagar for the comments on an early draft of this thesis, which significantly improved its contents and helped me to look further for the importance of some analysis. This work could not have happened without some people that not only helped me in fieldwork but also by making me feel it was not just work. These people were Joana Pereira, Cidália Gomes and Daniela Almeida. I would also like to thank the Molecular Evolution debate partners for the helpful discussions and all the interest shown in sharing experiences and solve problems.

To all my friends, thank you for the great interest and encouragement they have always shown. In particular I thank Tânia Nogueira for the assistance in matters of English syntax and grammar.

Finally, I would like to thank my close family who always supported me in all the academic decisions and for doing the possible and the impossible to provide the necessary conditions to finalize my academic degree. To them I dedicate the present thesis.

Abstract

Sensorial receptors are the first intermediaries between the environmental stimulus and the organism, being quite relevant study objects when used to understand the molecular basis of adaptation. Here, a photoreceptive sensorial gene, the melanopsin (*OPN4*), was studied by an integrated gene/protein evolutionary approach. Melanopsin is a photosensitive cell protein involved in the regulation of circadian rhythms and other non-visual responses to light. The melanopsin gene family is represented by two paralogs, *OPN4x* and *OPN4m*, which were originated through gene duplication early in the emergence of vertebrates. The aim of this work was to determine the molecular and evolutionary mechanisms that led to the adaptation of the melanopsin to different photic environments over time.

Our results showed two evolutionary phenomena that can be stated as particularly important to mediate the adaptive evolution of the melanopsin photoreceptors: (1) purifying or negative selection and (2) duplication events followed by minor episodes of positive selection. Firstly, we found evidence of positive selection and functional divergence in the sites near the retinal pocket that are likely related to the spectral sensitivity of melanopsins. This suggests that melanopsin *m* and *x* variants are sensible to different quantities or qualities of light. Additionally, we found evidence of destabilizing positive selection in the second and third intracellular loops. These loops are involved in establishing the activation of a specific G-protein type, and are therefore responsible for the specificity of the light signaling process. Since we found notorious amino acid variability in these loops, this may suggest some ambiguity in the G-protein interactions with the melanopsin. Our predictions showed that the Gio and Gq11 types are the most likely G-proteins mediating melanopsin phototransduction cascade.

Thus, we suggest that melanopsins paralogs (*OPN4m* and *OPN4x*) have diverged in function. On one hand, by specializing in different ranges of the light spectrum and, on the other hand, establishing ambiguous intracellular interactions with different G-protein types. Our results provide new insights on the phototransduction process, and additional tools for understanding the links between melanopsin gene evolution and the photic specializations observed in vertebrates. Moreover, the adaptive mechanisms that we have advanced not only allow us to understand how the signaling light pathways are performed in order to implement the regulation of circadian rhythms, but also to comprehend the adaptation to changing photic environments, as experienced by the vertebrates during their evolution.

Resumo

Os recetores sensoriais são os primeiros intermediários entre o estímulo ambiental e o organismo, sendo relevantes objetos de estudo para compreender as bases moleculares da adaptação. Neste estudo, um gene sensorial fotorecetivo, a melanopsina (*OPN4*), foi estudado numa abordagem evolutiva integrada (gene e proteína). A melanopsina é uma proteína membranar fotorecetiva envolvida na regulação dos ritmos circadianos e noutras respostas não-visuais à luz. Existem dois genes parálogos (*OPN4x* e *OPN4m*) a representar a família das melanopsinas que surgiram por duplicação durante o aparecimento dos vertebrados. O objetivo deste trabalho foi determinar que mecanismos moleculares e evolutivos permitiram a adaptação das melanopsinas nos diferentes ambientes fóticos.

Os nossos resultados mostraram que dois fenómenos evolutivos foram importantes durante a evolução das melanopsinas: (1) a seleção negativa ou purificante e (2) os eventos de duplicação, seguidos de pequenos episódios de seleção positiva. Primeiramente, encontrámos evidência de seleção positiva e divergência funcional nos resíduos responsáveis pela acomodação do retinal. Tal sugere que as variantes *m* e *x* da melanopsina deverão ser sensíveis a diferentes quantidades ou qualidades de luz. Adicionalmente, encontrámos evidência de seleção positiva destabilizante nas segunda e terceira loops intracelulares. Estas loops estão envolvidas na ativação de proteínas G, sendo por isso responsáveis pela especificidade do processo de sinalização da luz. A notória variabilidade aminoacídica encontrada nestas loops, sugere que alguma ambiguidade deverá existir na interação das proteínas G com a melanopsina. De facto, as nossas previsões mostraram que dois tipos de proteínas G (Gio e Gq11) devem mediar a cascata de fototransdução na melanopsina.

Assim, sugerimos que os parálogos *OPN4m* e *OPN4x* divergiram em função, por um lado especializando-se na sinalização de diferentes qualidade de luz, e por outro, estabelecendo interações ambíguas com diferentes tipos de proteínas G. Os nossos resultados ajudam a perceber melhor o processo de fototransdução das melanopsinas, bem como fornece ferramentas adicionais para entender as relações entre a evolução das melanopsinas e as especializações observadas ao nível fótico nos vertebrados. Assim, os mecanismos adaptativos que avançamos, ajudam, por uma lado, a perceber melhor como as cascatas de transdução do sinal luminoso acontecem para levarem a cabo a regulação dos ritmos circadianos, e por outro, explicam como a é feita a adaptação dos fotorreceptores dos vertebrados em diferentes e dinâmicos ambientes fóticos durante a sua evolução.

Index

List of Figures

1. General introduction

The evolution of the eye has always intrigued those who were devoted to its study and comprehension, not only because of its admirable anatomical complexity, but also because it is an interesting example of sensorial adaptation in the animal kingdom. Charles Darwin himself has considered the evolution of the eye and its adjacent structures as a complex and imaginative challenge. He refers in **On the Origin of Species by means of Natural Selection** [1],

> *To suppose that the eye, with all its inimitable contrivances* [...] *could have been formed by natural selection, seems, I freely confess, absurd in the highest possible degree* [...] *Yet reason tells me, that if numerous gradations from a perfect and complex eye to one very imperfect and simple, each grade being useful to its possessor, can be shown to exist* [...] *and if any variation or modification in the organ be ever useful to an animal under changing conditions of life, then the difficulty of believing that a perfect and complex eye could be formed by natural selection, though insuperable by our imagination, can hardly be considered real.*
>
> Charles Darwin, On the Origin of Species by means of Natural Selection, 1907

Looking further in the diversity of the sensorial adaptations in the animal kingdom we can conclude that most animals have some kind of light sensibility. During evolution, independent ways of signalling light were developed in the animal lineages, from simple features, as the planarian eye spot, to more complex structures as the chambered eye-type of cephalopods [2]. It is believed that the way organisms signalize light was not immune to the specific photic environments that animals occupy. In fact the visual capacity is involved in such basilar processes, as the circadian rhythm regulation and the identification of possible preys and predators and it is expected that the visual systems have modified to supply the specific photic conditions required for where the animals live [3, 4]. Consequently, the visual systems are one of the best examples of notorious ecological adaptation to specific niches and also one of the best sensorial system studied. We can refer some examples (**figure 1.1**):

- The rudimentary eyes of the mole rat (*Spalax ehrenbergi*) are located under the skin and do not respond to light stimuli [5]. However, the removal of the eyes disturbs photoperiod perception in these animals. This kind of eye, that does no longer implement visual functions, act as an optical control to adjust the life cycle of *Spalax*. The visual stimuli help to determine whether it is night or day, summer or winter which is essential to maintain body temperature, muscles activity and to decide if it is time for mating [6].

Figure 1.1. Visual adaptations in the animal kingdom. Three examples of special photic adaptations in vertebrate group: **A.** the rudimentary eyes of the mole rat, (*Spalax ehrenbergi*), **B.** plumage patterns of the blue tit (*Parus caeruleus ultramarinus*) and **C.** the four-eyed fish (*Anableps anableps*).

- Birds are very complex in their visual acuity. Some birds are sensible to ultraviolet (UV) light, and its perception is thought to be involved in sexual dimorphism and prey identification [7]. There are plumage patterns signalized in ultraviolet light completely invisible to the human eye, making impossible to us to discriminate which is the male or female. However, they can be distinguished by the presence of ultraviolet reflective patches on their feathers. Recent work has shown that ultraviolet plumage reflectance is an important signal used by blue tit (*Parus caeruleus ultramarinus*) females during mate

choice. This proves the importance of non-visual spectral zones mediating bird sexual behavior [8, 9].

- Water environments are very complex on the photic point of view, comprising so many factors including turbidity, salinity, pressure and depth, which determines very different refractive indexes on the water column [10]. As expected fish show unusual light-related adaptations. The four-eyed fish (*Anableps anableps*) is able to see simultaneously in air and water; it normally swims so that the eye is bisected by the water meniscus. Each eye has a single retina, but the dorsal part receives images from the air, and the ventral region receives aquatic images [11, 12].

1.1. Vertebrate photoreceptors

The capacity to signalize light is very important to animals because it provides many advantages: identify suitable prey and detect potential predators, control diurnal cycles and seasonal changes in day length and receive information about periods of feeding, reproduction and location of shelter [13]. The way organisms signalize light has profound effects on the organism fitness, since this sensorial capacity is involved in many essential processes we can admit that animals modify their visual systems according to their special photic environments [14]. During the evolution of vertebrates, sensorial receptors, including photoreceptors, showed an enormous specialization and complexity, which can be explained by adaptive radiation through the colonization of many different environments [3, 4, 14]. The eye is the most studied photoreceptive organ in vertebrates, but there are also others, called the non-mammalian photoreception organs. They are the pineal gland, deep brain, peripheral oscillators and dermal tissue [15] (**figure 1.2**). Indeed, comparing all vertebrates, mammals seem to be the most simple organism from a visual point of view, showing lower number and less complexity on their photoreceptive organs [16].

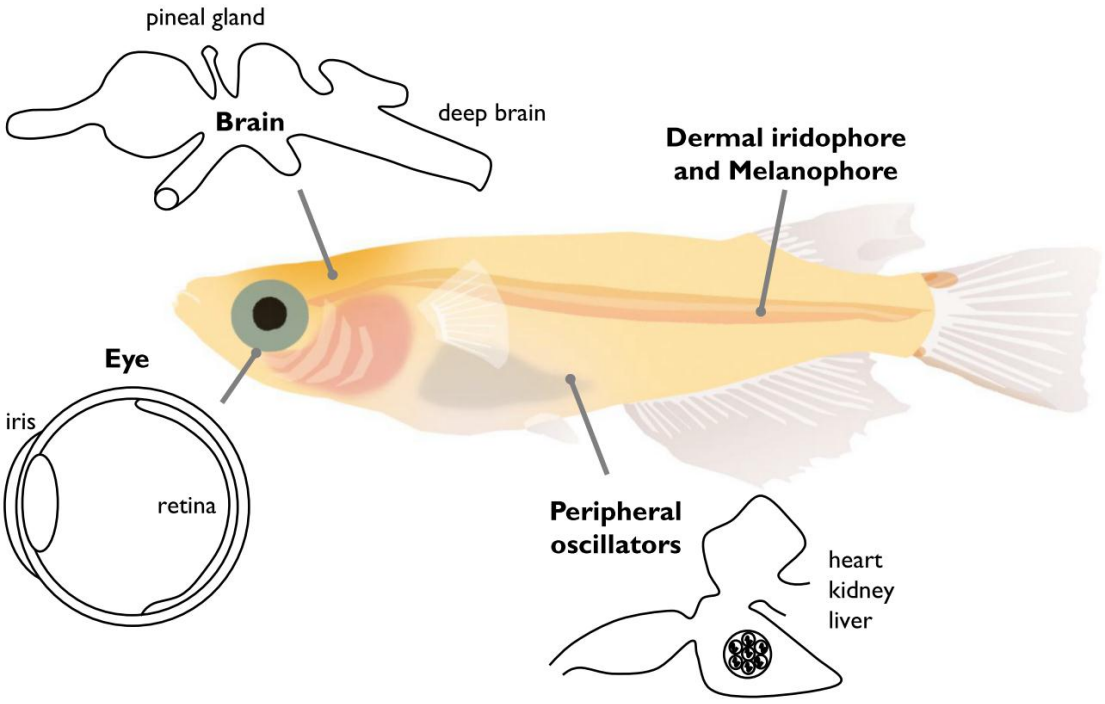


Figure 1.2. Photoreception organs vertebrates. Similarly to the classical photoreceptors within the retina of the lateral eye, direct photoreception in the isolated iris has also been described. In non-mammalian species the pineal complex contains photoreceptors as well as the deep brain. Dermal photoreception has been described in amphibians and fish. Moreover, the zebrafish peripheral tissues have been shown to be able to entrain their molecular oscillators directly to light. Adapted from [15].

Some photoreceptors performed visual functions as the identification of colors and the light intensity and/or brightness, but others can also perform non-visual, but light-dependent functions, as the regulation of circadian rhythms, control of the internal temperature and the pupillary light reflex [15].

1.2. Diversity of opsin photopigments

Photoreception organs contain photoreceptors, which are types of cells that are composed by one or several photoreceptive molecules. These photoreceptive molecules are called opsins and are members of the G-protein-coupled receptors (GPCRs) protein group. GPCRs strucutre is characterized by seven transmembrane domains, three intracellular and extracellular loops and the carboxyl (C) and the amine-terminus (N-terminus) [17, 18]. The external loops recurrently establish the connection with the ligand, and the internal ones establish the connection with the G-proteins (guanine nucleotide-binding proteins), that are responsible for the transduction of the signal in the cell interior [19]. GPCRs are also represented in other

sensorial molecules as the chemosensory receptors involved in the sense of smell and taste [20, 21].

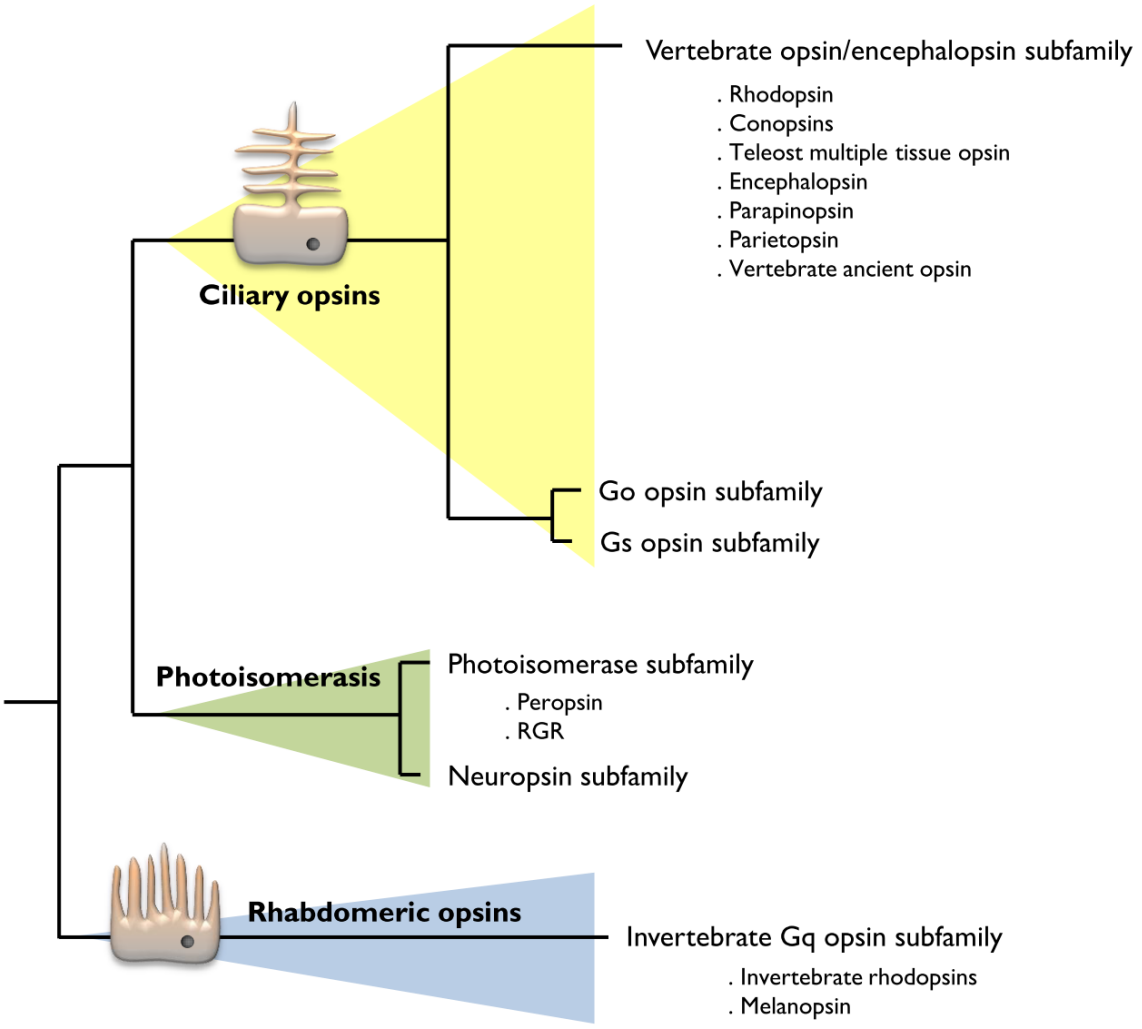


Figure 1.3. Schematic phylogenetic tree of the opsin family. There are three groups: ciliary-opsins and rhabdomeric-opsins. Additionally, a non-defined miscellaneous group more closely related to c-opsins can be also referred, which includes the photoisomerasis opsins. The ciliary and rhabdomeric types of opsins were likely to diverge before bilaterians emerged. Adapted from [22, 23].

Opsins act as pigments that activate G-proteins in a light-dependent manner. All opsins mediate the light transduction in three steps: (1) during light detection the retinal chromophore absorbs a photon and its photoisomerization leads to an all-trans state; (2) this implies a conformational change in the retinal that allows the opsin to interact with a G-protein; and (3) the G-protein activates the phototransduction cascade which finally changes the receptor membrane potential [24, 25]. Opsins are then the first proteins on the phototransduction

cascade mediating the environmental photic signals. As we may expect they show photic specific sensitivity, which means that different opsins can be sensible to different quality and quantity of light [26]. Most of the opsins have their spectral maximum sensitivity in the visual light spectrum, however it was already reported photoreceptors that are sensible to infrareds as well as UV light [27].

Most opsins belong to two major groups: (1) the rhabdomeric type (r-opsins) which is characterized by apical microvilli topology, the activation of the phosphoinositol signalling cascade (Gq) and membrane depolarization in response to light; (2) the ciliary type (c-opsins) which is characterized by extended membranes to form a modified cilium topology, cyclic nucleotide signalling cascades (Gs, Gt and Go) and hyperpolarization in response to light [23, 28] (**figure 1.3**). Vertebrate and invertebrate photoreception organs can show both ciliary and rhabdomeric types of opsins, but ciliary types are more common in vertebrates and rhabdomeric in invertebrates [29]. Current evidences show that these major groups of opsins diversified before the deuterostomes split from the protostomes [18].

Based on sequence homology and G-protein preferences, the opsin family can be characterized into six subfamilies, namely the vertebrate opsin/encephalopsin subfamily, the Go opsin subfamily, the recently characterized Gs opsin subfamily, the invertebrate Gq opsin subfamily (that includes the melanopsin group), the photoisomerase subfamily and the neuropsin subfamily [17, 22, 30] (**figure 1.3**).

1.3. Melanopsins

Melanopsins (OPN4) are photopigments found in the vertebrate eye, more specifically, in the specialized photosensitive ganglion cells (iPGCs) in the inner retina of mammals [31]. Melanopsin is responsible for the regulation of the circadian rhythms and pupillary light reflex [32], but it is also involved in the light suppression of activity, alertness and acute suppression of pineal melatonin [33, 34]. As photoreceptors, melanopsins are more sensible to blue light, and their photosensitive response is selectively sensitive to short-wavelength light (approximately 480 nm) [35–37].

Melanopsins were discovered by Ignacio Provencio and his colleagues in the dermal melanophores of *Xenopus laevis* [38]. In 1999, Russel G. Foster was able to show that despite the absence of the rhodopsin and conopsin visual pigments in the eye, the entrainment of mice to a light-dark cycle was maintained, suggesting that a third photopigment must exist in the mammalian eye [39]. Three years later, Hattar and his colleagues showed that melanopsins are involved in the entrainment of the suprachiasmatic nuclei (SCN), responsible for controlling circadian rhythms in mammals [32].

In vertebrates, melanopsin is expressed in the retina only in 1 to 2% of the ganglion cells [40–42] and their response to single photon is more prolonged and large when compared to rhodopsin light response [43]. Moreover, melanopsin amino acid sequence resembles the invertebrate visual photopigments [44, 45], as well as, the signalling cascade they are involved in. Indeed, the melanopsin phototransduction mechanism is not clearly understood. The evolutionary relationship between melanopsins and invertebrate opsins allows to conclude some connection with melanopsin and invertebrate signal transduction pathways [46]. In this case, melanopsins should interact with a Gq/11 which activates PLC (phospholipase C) that ultimately modulate TRPC (transient receptor protein channel) channel giving rise to cellular depolarization [47–49]. Recently, a Gq/11-triggered PLC light-signalling cascade was described in amphioxus [50].

Melanopsins show seven helical transmembrane domains (TM) with two C and N terminus and three intracellular (IC) and extracellular loops (EC). Based on the rhodopsin secondary and crystal structure it is possible to infer several motifs with functional and structural relevance for melanopsins [51, 52] (**figure 1.4**): two cysteine residues in the TM3 and EC2 domains that are involved in disulfide bond formation; a tyrosine and a glutamic acid in the TM3 and EC3 domains, respectively, which act as counter ions to the positive charge of the protonated Schiff base; a DRY motif at the TM3/EC2 boundary that provides a negative charge to stabilize the inactive opsin molecule; a lysine residue in the TM7 domain that is covalently linked to the retinal chromophore; and a conserved NPxxY(x)2,3HPKF (named NP-Y-F) motif in the TM7-H8 region conferring structural integrity upon pigment activation.

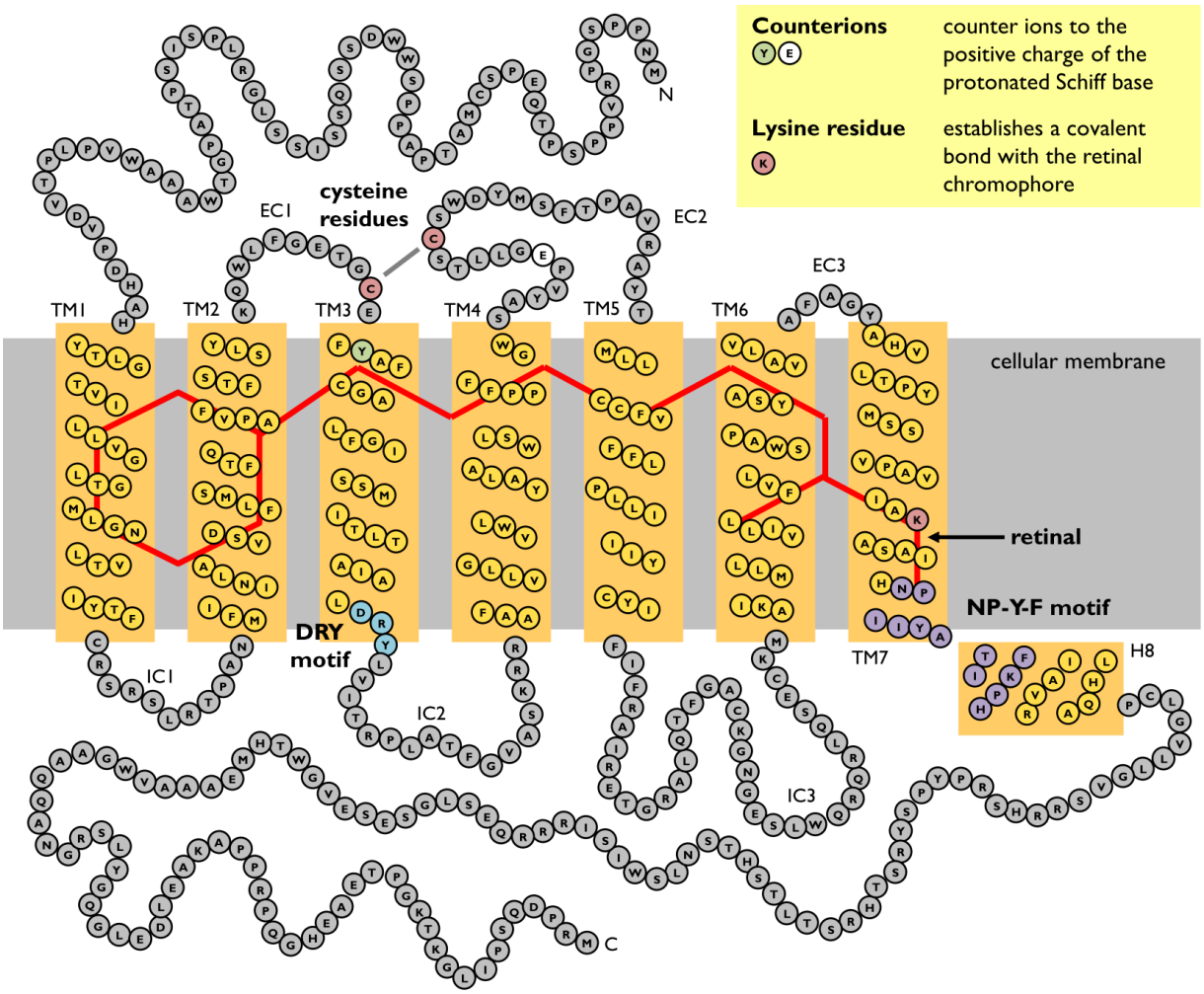


Figure 1.4. Schematic representation of the human melanopsin. (variant 1, NP150598) showing the presence of 7 helical transmembrane (TM) domains (TM1-7), an amino-terminus (N), 3 extracellular (EC1-3) and 3 intracellular loops (IC1-3), a C-terminus, a predicted N-terminus and a putative eighth cytoplasmic helix (H8). Adapted from [51, 52].

Melanopsin has been described in all deuterostome classes, including the urochordates, cephalochordates and all vertebrate genomes [53]. Particularly in vertebrates, an additional gene copy is found, from which two melanopsin paralogs were described so far: the mammalian type melanopsin (*OPN4m*) and the *Xenopus*-type melanopsin (*OPN4x*). The *OPN4x* paralog is not present in mammals, suggesting that this variant was lost during the mammalian emergence [54]. Recently, numerous melanopsin variants were described in teleost fish including *OPN4x1*, *OPN4x2*, *OPN4m1*, *OPN4m2* and *OPN4m3* [55].

It is not completely understood which function each paralog performs in the visual systems of vertebrates. So far, we know that the *OPN4m* variant is widely expressed in a subset of photoreceptive retinal ganglion cells of the eye [56], while the non-mammalian vertebrates also express it on the intraocular photoreceptors such as the pineal gland and deep brain [57, 58]. Additionally, *Opn4x* appears to have a much weaker light-induced response and a confined expression in the extra-ocular photoreceptors [54].

2. Motivation and Objectives

As we can see, the study of photoreceptor types, their morphological characteristics and origins is very important to disentangle the way animals integrate the visual stimuli. Our interest was to understand the molecular adaptation in photoreceptors. Since phototransduction signalling and regulation in vertebrates is a very complex task, the evolutionary study of the melanopsin gene family would increase our comprehension of the evolution of vertebrate circadian rhythm regulation. Moreover, it would provide insights on the molecular basis of the visual adaptation of photoreceptors during vertebrate evolution.

The main motivational reasons were:

- Photoreceptors are opsins that are expressed by protein-coding genes involved in the eye development [59, 60]. Opsin activity and the related biochemical cascade are highly described in the literature [24, 25]. Additionally, the structural and physiological characteristics of opsins are highly conserved which allows making global considerations [17, 18].
- The genomic and proteomic evolutionary approach was never applied to the specific group of the melanopsin subfamily. Thus, this justifies a deeper study in this protein group which could be very important for understanding their relationship with the circadian rhythms.
- Furthermore, there is enough genetic data available in the public databases such as NCBI and Ensembl [61, 62]. The vertebrate's group counts with sixty-nine genome projects represented mostly by mammals, but also by teleost fish, amphibians, reptiles and birds.

The main objective of this study was **to assess the evolutionary history and adaptive evolution patterns, at the gene and protein level, of the melanopsin gene family (*OPN4m* and *OPN4x* paralogs) in vertebrates**. However, other specific objectives can be stated:

1. Determine the main events of gene gain and loss in the melanopsin gene family that led to the current gene content in the major groups of vertebrates. Then, infer possible environmental reasons for the loss and gain of genes in particular vertebrate lineages.

2. Analyze the synteny patterns of the *OPN4m* and O*PN4x* neighbor genes and determine the possible relationships with the previously described duplication events in vertebrates – the 2R and 3R events [63, 64].

3. Study the selective pressures following the duplication events at the branch level, especially the one that gave rise to the *OPN4m* and *OPN4x* paralogs, as well as to indicate the influence of neo-functionalization and/or sub-functionalization processes after duplication.

4. Investigate the directionality of the selective pressures in the melanopsin molecule residues and infer possible domains of fast or slow evolution. Relate these domains to their location in the three-dimensional structure of the protein and establish possible functional or structural implications.

5. Describe the conservation and the amino acid patterns in the sites involved in the isomerization of the 11-cis-retinal, the second and third intracellular loops. It is intended to understand the adaptive potential of these sites in signalling light, as well as possible interactions at intracellular level in the photoreceptive cascade.

Altogether, these objectives intend to provide new insights on the phototransduction process and additional tools for disentangling the links between melanopsin gene evolution and the circadian rhythms specializations observed in vertebrates.

3. Prior methodological considerations

This section aims to address key concepts discussed in the field of molecular evolution and to present some possible definitions of the main phenomena that are currently thought to be involved in the evolution of genomes, genes and proteins. In addition to the theoretical approach, it is intended to introduce the state-of-the-art assumptions and ideas that are currently widely discussed.

3.1. Sequence evolution after gene duplication

The evolution of genes is typically conservative in the absence of gene duplication [65], according to Ohno's classic view. Duplication of genetic material (including regulatory elements) is generally accepted to play an important role in generating the raw material to be subject to natural selection [66, 67] (**figure 3.1**). Tandem, linear or whole-genome duplication (WGD) events produce pairs of similar genes, which can ultimately become differentiated throughout substitutions, selection and/or syntenic rearrangements [68].

The most likely fate of a duplicated gene is to degenerate into a pseudogene by integrating non-sense mutations, or be lost from the genome due to chromosomal reorganization and partial or complete deletion [69]. Right following the duplication event, one copy is under purifying selection and the other copy is free to accumulate neutral or nearly neutral mutations [70]. Gene loss through the non-functionalization is a high frequent process because only one copy is required to maintain the ancestral function. A less frequently, but expected outcome is when a new beneficial variant is acquired, as the result of advantageous substitutions in coding or regulatory regions. Classical models predict that few duplicates should be retained in the genome during time, hence neo-functionalization events must be extremely rare [70, 71].

Experimental evidence from the expression analyses suggest not only that extant paralog genes could have segregated the multiple functions of their single ancestral gene, but also that multi-function genes are more likely to generate gene families by retained duplicated copies in the genome [70]. Force et al. 1999 introduced the sub-functionalization model [72, 73] proposing that after gene duplication, both copies are necessary to implement the functions

that was assured by the original ancestral gene. Therefore, some types of genes have structural and functional features that allow them to adapt easily the segregated functions while others, probably because gene duplication has an intermediate fitness-decreasing effect, are less predisposed to experience functional or structural innovation [74, 75]. Genomes include gene families with great number of gene representatives, likely to be generated by sub-functionalization (opsin photoreceptive molecules or the odorant chemoreceptor families) but how genes evolve the multiple functions in the first place is a handicap of the sub-functionalization model [76].

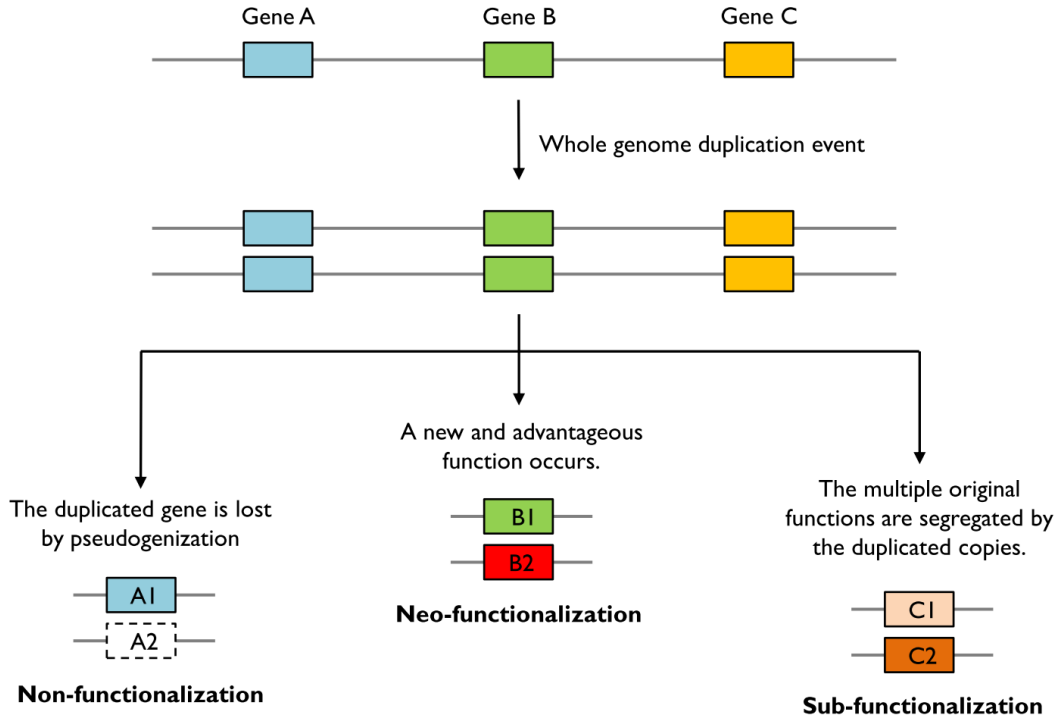


Figure 3.1. Evolutionary fate of duplicated genes. Following the whole genome duplication events each copy can have different fates: gene A2 copy degenerates to a pseudogene by chromosomal reorganization, partial or complete deletion or non-sense mutations; gene B2 copy acquires a new advantageous substitution in the coding or regulatory regions; genes C1 and C2 are both necessary to implement the ancestral function of gene C. Adapted from [72].

In chordates, it is well determined the main whole genome duplication events that happened along their evolutionary history. They contributed to the large number of paralog genes and gene families that are present in chordate genomes, especially in the vertebrates [77, 78]. Onho advocates that two rounds of WGD occurred between the origin of chordates and the origin of jawed vertebrates (1R and 2R) [79]. The 2R duplication event was particularly

important for the vertebrates, being responsible for their anatomical and physiological differentiation (e.g. the complex eye) when compared with their closest living relatives, the urochordates and cephalochordates. The 1R and 2R whole genome duplication events seem to be the likely explanation for all the morphological diversity that we can currently see in vertebrates [63]. Whole genome duplication events shaping the genomes of vertebrates have not only been proposed in the early evolution of vertebrates, but also in the stem lineage of actinopterygian fishes, after their divergence from the land vertebrates [64]. Indeed, the gene content of most of the gene families in vertebrates is extensively represented by a great number of duplicated copies in the teleost fish genomes [80].

3.2. Positive, negative and neutral selection

Adaptation by natural selection is a very important process in evolution, explaining the incredible complexity and diversity of proteins, cells and organisms. The adaptive process implies, at the molecular level, a dynamical change of the genetic material and consequently, the appearance of novelties [81, 82]. These novelties can pass through generations because they represent an advantageous feature to the organism survival and reproduction [83] and ultimately can be extended to all population.

Mutation is the first source of all genetic variation, being responsible for producing new alleles in the population. However, this evolutionary force itself is not capable of changing the allele frequencies in the population [84]. The fate of a new polymorphism produced by mutation can then be driven by natural selection or genetic drift which differently from mutations can change the allele frequency in the population. Natural selection can act in two directions, increasing the frequency of a beneficial allele until it is fixed in the population or decreasing the frequency of a deleterious allele until it is eliminated [85, 86]. Genetic drift act fixing or eliminating randomly the neutral mutations from the population (**figure 3.2A**) [84, 86]. Altogether, the nucleotide and/or amino acid patterns produced by these evolutionary forces help to clarify the evolutionary history and the episodes of adaptation that occurred in the past.

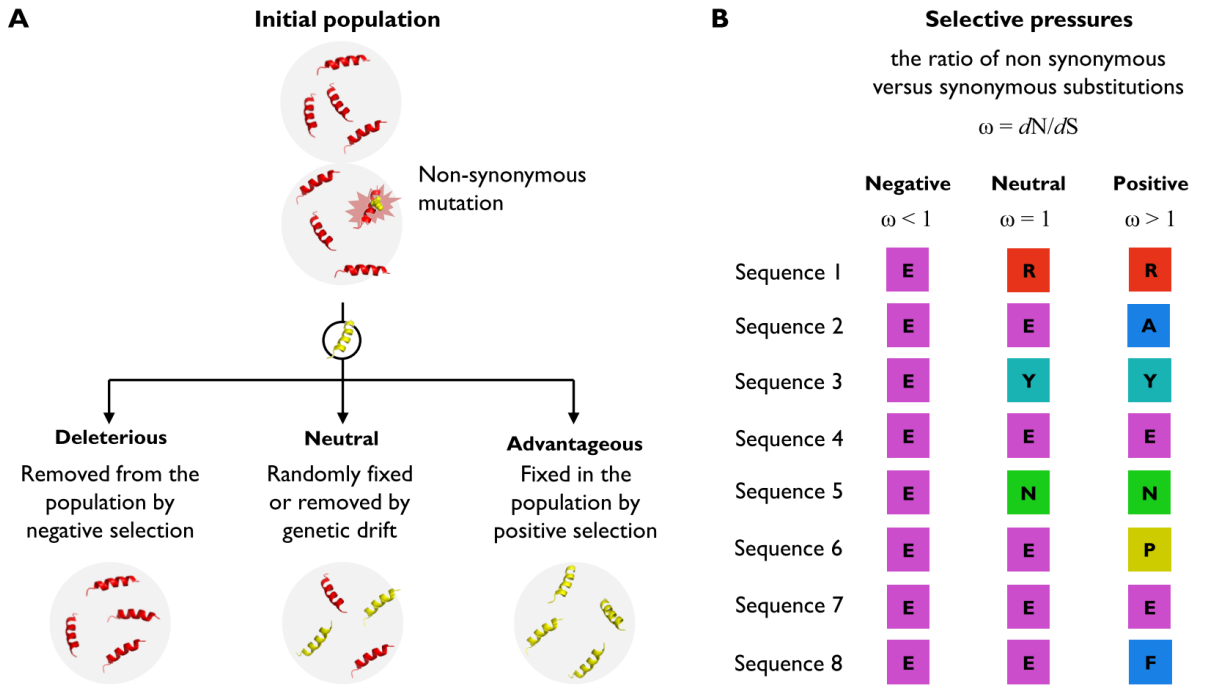


Figure 3.2. Natural selection and genetic drift evolutionary forces. A. The fate (fixation or elimination) of a new polymorphism in a population whether it is deleterious, neutral or advantageous. B. Amino acid patterns representative of the neutral, negative and positive selective signatures in a multiple sequence alignment.

Within the coding regions, substitutions can be non-synonymous if it implies a change in the amino acid specified by a codon, or synonymous if the substitution do not change the amino acid specified by the codon, due to the redundancy of the genetic code. In order to assess the selective signatures of the positive, negative and neutral selection the comparison of non-synonymous (dN) and synonymous (dS) substitutions rates is frequently used. The ratio between these two rates, measured as $\omega = d\text{N}/d\text{S}$, reflects the effect of selection on the protein-coding genes [87, 88]. Thus, if non-synonymous substitutions are deleterious, purifying or negative selection will prevent the fixation of the polymorphism ($d\text{N} < d\text{S}$) and the ω-ratio is expectedly lesser than 1, whereas if non-synonymous substitutions are neutral then they can be fixed as probable as the synonymous substitutions ($d\text{N} \approx d\text{S}$) and the ω-ratio is expectedly equal to 1. Under positive or diversifying selection the ω-ratio is expectedly higher than 1. In this case non-synonymous substitutions represent an adaptive advantage, and consequently they become fixed in the population much more frequently than synonymous substitutions ($d\text{N} > d\text{S}$) (**figure 3.2B**) [89, 90].

According to the Kimura's Neutral Theory most of the observed polymorphisms, both within species and between species, is due to random fixation of selectively neutral substitutions [87]. Thus, to test adaptive evolution, the hypothesis of neutral evolution must be rejected, meaning that the ω-value must be significantly greater than 1 [90, 91]. Codon substitution models assume different ω-values for all branches of a phylogeny, allowing testing changes in selective pressures following clade divergence [90, 92]. Other codon models, known as site-models, allow the ω-ratio to vary among amino acid sites [93–95]. These are particularly useful to infer important structural or functional sites and domains along the protein, by recognizing the fast and slow-evolving regions. A third type of model, the branch-site models, account simultaneously for variation in the selective forces among sites and branches [96].

3.3. Destabilizing positive selection and functional divergence

Models of selection that exclusively use the ω-ratio statistics to detect positive selection are generally not sensitive enough to detect minor episodes of molecular adaptation [97, 98]. Firstly, it is not possible to conclude that positive selection episode has not occurred if ω is not statistically higher than 1, since single amino acid changes can be adaptive if a certain biochemical property is taken in consideration. Otherwise, it is not possible to conclude an episode of positive selection if the ω-ratio is statistically significant higher than 1 as non-synonymous substitutions can represent different amino acids but with similar biochemical properties. Thus, a new family of methods recently has been proposed that approaches the question of amino acid replacement diagnosis using changes in physicochemical amino acid properties [98].

The destabilizing positive selection concept has been successfully used in a variety of protein studies for identifying and characterizing adaptation in terms of shifts in the physicochemical properties of amino acid replacements [99]. It differentiates adaptive replacements from those that may have resulted from random mutation. Destabilizing positive selection measures how destabilizing a non-synonymous substitutions can be taking into account a specific amino acid property (e.g. hydrophobicity, isoelectric point) [100]. In this perspective, it is also possible to asses stabilizing positive selection when the observed non-synonymous substitutions do not represent a notorious change at the biochemical level (**figure 3.3A**). Adaptive changes may

then be mapped onto three–dimensional structures to qualitatively assess the degree to which adaptive changes are associated with functional domains and motifs [101, 102].

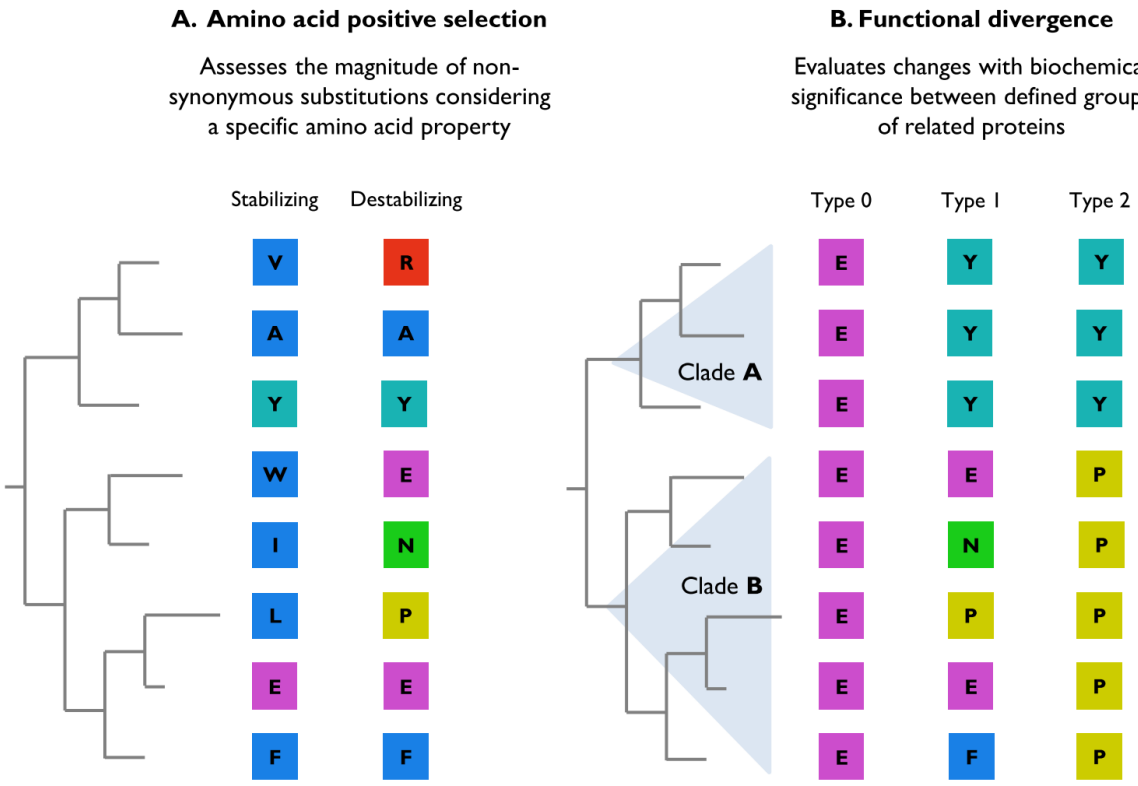


Figure 3.3. Amino acid configuration diagnosis evaluating physicochemical properties. A. Destabilizing and stabilizing positive selection measure the non-synonymous variation considering the magnitude of the substitution for a specific amino acid property. B. Functional divergence evaluates type I and type II amino acid configurations in order to assess the involvement of differentiated constrained evolution and/or radical change in amino acid properties between two defined clades.

Gene function can be assessed quantitatively at the site level considering the functional constraints of the protein sequence. Therefore, an amino acid residue is said to be functionally important if it is evolutionarily conserved and thus, changes to that conservation status may indicate the contribution of functional divergence [103, 104]. We have seen that following gene duplication events, gene copies can suffer specialization and so it is desirable to identify amino acid sites that are responsible for the functional diversity in the respective paralog gene families [105, 106].

According to observed amino acid configurations in the multiple sequence alignment, three basic types of functional divergence can be characterized [104–106]: type 0 represents an amino acid configuration in which the site is fully conserved in all the paralog genes, suggesting that these residue is important for the shared function of the respective gene family; type I represents amino acid configurations that are highly conserved in one clade but variable in other clade, denoting that these residues have experienced differentiated functional constraints in a particular site; and type II represents amino acid configurations that are very conserved in both clades but the amino acid is different between clades, implying that these residues may be responsible for functional specification specially when the substitution has some biochemical significance [105]. According to these amino acid configurations it is possible to determine two basic types of functional divergence, the type I functional divergence represent altered functional constraints between clades and the type II functional divergence represent a radical change in amino acid properties between clades but no altered evolutionary rates [105] (**figure 3.3B**).

Most amino acid changes occurs neutrally and do not represent signatures of functional divergence, thus it is important to develop suitable statistics to distinguish type I and II functional divergence from nearly random amino acid patterns [107]. In 2002, Gu introduced a fundamental statistic for functional divergence after gene duplication, the coefficient of functional divergence (θ), expressing the rate correlation between two duplicated genes. For instance, if the site is evolving equally in both clades, as it occurs in the type 0 amino acid configuration, the coefficient of functional divergence is 0. Whereas, if the site is evolving differently in between clades, as the type I and II amino acid configurations, the coefficient of functional divergence is 1 [104–106].

4. The role of gene duplication and unconstrained selective pressures in the melanopsin gene family evolution and vertebrate circadian rhythm regulation

The following manuscript was submitted and accepted for publication.

OPEN ACCESS Freely available online

PLOS ONE

The Role of Gene Duplication and Unconstrained Selective Pressures in the Melanopsin Gene Family Evolution and Vertebrate Circadian Rhythm Regulation

Rui Borges[1,2], Warren E. Johnson[3], Stephen J. O'Brien[3,4], Vitor Vasconcelos[1,2], Agostinho Antunes[1,2,3]*

1 CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas, Porto, Portugal, **2** Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, Porto, Portugal, **3** Laboratory of Genomic Diversity, National Cancer Institute, Frederick, Maryland, United States of America, **4** Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg, Russia

Abstract

Melanopsin is a photosensitive cell protein involved in regulating circadian rhythms and other non-visual responses to light. The melanopsin gene family is represented by two paralogs, *OPN4x* and *OPN4m*, which originated through gene duplication early in the emergence of vertebrates. Here we studied the melanopsin gene family using an integrated gene/protein evolutionary approach, which revealed that the rhabdomeric urbilaterian ancestor had the same amino acid patterns (DRY motif and the Y and E conterions) as extant vertebrate species, suggesting that the mechanism for light detection and regulation is similar to rhabdomeric rhodopsins. Both *OPN4m* and *OPN4x* paralogs are found in vertebrate genomic paralogons, suggesting that they diverged following this duplication event about 600 million years ago, when the complex eye emerged in the vertebrate ancestor. Melanopsins generally evolved under negative selection ($\omega = 0.171$) with some minor episodes of positive selection (proportion of sites = 25%) and functional divergence ($\theta_I = 0.349$ and $\theta_{II} = 0.126$). The OPN4m and OPN4x melanopsin paralogs show evidence of spectral divergence at sites likely involved in melanopsin light absorbance (200F, 273S and 276A). Also, following the teleost lineage-specific whole genome duplication (3R) that prompted the teleost fish radiation, type I divergence ($\theta_I = 0.181$) and positive selection (affecting 11% of sites) contributed to amino acid variability that we related with the photo-activation stability of melanopsin. The melanopsin intracellular regions had unexpectedly high variability in their coupling specificity of G-proteins and we propose that Gq/11 and Gi/o are the two G-proteins most-likely to mediate the melanopsin phototransduction pathway. The selection signatures were mainly observed on retinal-related sites and the third and second intracellular loops, demonstrating the physiological plasticity of the melanopsin protein group. Our results provide new insights on the phototransduction process and additional tools for disentangling and understanding the links between melanopsin gene evolution and the specializations observed in vertebrates, especially in teleost fish.

Citation: Borges R, Johnson WE, O'Brien SJ, Vasconcelos V, Antunes A (2012) The Role of Gene Duplication and Unconstrained Selective Pressures in the Melanopsin Gene Family Evolution and Vertebrate Circadian Rhythm Regulation. PLoS ONE 7(12): e52413. doi:10.1371/journal.pone.0052413

Editor: Stephen R. Proulx, University of California Santa Barbara, United States of America

Received July 25, 2012; **Accepted** November 15, 2012; **Published** December 21, 2012

Funding: The authors acknowledge the Portuguese Foundation for Science and Technology (FCT) for financial support to RB (SFRH/BD/79850/2011) and the projects PTDC/AAC-AMB/104983/2008 (FCOMP-01-0124-FEDER-008610) and PTDC/AAC-AMB/121301/2010 (FCOMP-01-0124-FEDER-019490) to AA. This research received support by the Intramural Research Program of the National Institutes of Health, National Cancer Institute to WEJ and SJO. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: aantunes@ciimar.up.pt

Introduction

Vertebrates have a wide range of strategies to respond to light in different photic environments [1]. The evolution of these diverse light-signalling processes and the link between photoreceptors and adaptive strategies are not fully understood. One of the most-recently discovered groups of photoreceptors, melanopsin (*OPN4*), was first described in the dermal melanophores of *Xenopus laevis* [2]. Its main functions are non-image forming, including the regulation of circadian rhythms, the pupillary light reflex and melatonin synthesis [3–5]. Melanopsins are sensitive to low wavelength light with maximum sensitivities near to 480 *nm* [6,7].

Within vertebrate genomes there are two variants of the melanopsin gene: the mammalian-like melanopsin (*OPN4m*) and the *Xenopus*-like melanopsin (*OPN4x*) [8]. In mammals, only the *OPN4m* gene has been described, suggesting that the *OPN4x* variant was lost during mammalian evolution [9]. Mammalian melanopsin is expressed in a subset of intrinsically photosensitive retinal ganglion cells (ipRGCs) of the eye [10] while the non-mammalian vertebrates also express melanopsin in intraocular photoreceptors such as the pineal gland and deep brain [11,12]. Recently, numerous melanopsins were describe in teleost fish including *OPN4x1*, *OPN4x2*, *OPN4m1*, *OPN4m2* and *OPN4m3* [13].

Melanopsins are members of the G protein-coupled receptor (GPCR) protein family that is characterized by a heptahelical transmembrane conserved structure and the activation of a G-protein in their signalling transduction pathway [14]. Melanopsin structure includes seven helical transmembrane domains (TD), three intracellular (IL) and three extracellular (EL) loops, eight cytoplasmic domain (CD8), and N and C-terminals [15]. Residues that are critical for correct melanopsin conformation include: (i) two cysteine residues in the TD3 and EL2 domains that are

involved in disulfide bond formation, (ii) a tyrosine and a glutamic acid in the TD3 and EL3 domains, respectively, that act as counter ions to the positive charge of the protonated Schiff base, (iii) a DRY motif at the TD3/EL2 boundary that provides a negative charge to stabilize the inactive opsin molecule, (iv) a lysine residue in the TD7 domain that is covalently linked to the retinal chromophore, and (v) a conserved NPxxY(x)2,3HPKF (NP-Y-F) motif in the TD7-CD8 region conferring structural integrity upon pigment activation [15,16].

Koyanagi et al. proposed that rhabdomeric opsins evolved in protostomes to provide visual functions (*InRHO*) and in deuterostomes to provide non-visual functions (*OPN4*) [17]. It is hypothesized that all rhabdomeric photoreceptor share the same signal transduction pathway, including the activation of phospholipase C (PLC) and the inositol phosphate (IP3) pathway, which involves the Gq/11 G-protein type [18]. There are three families that constitute the major functional classes of G proteins and that are likely to mediate the melanopsin phototransduction cascade. The Gs and the Gi/o classes of G-proteins mediate the opposing effects of stimulation and inhibition of adenylate cyclase activity, and the Gq/11 family activates phospholipase C enzymes, resulting in phosphatilinositol hydrolysis [19]. Recently, a Gq11-triggered PLC light-signalling cascade was described in amphioxus [20], but a general model for vertebrate melanopsin phototransduction pathway is still missing. However, expression patterns in heterologously [18,21] and cultured melanophores and ipRGCs cells [22,23] strongly suggest the involvement of a Gq–based pathway.

Since the regulation of phototransduction in vertebrates is a very complex task, the study of the melanopsin gene family would increase our understanding of the evolution of vertebrate circadian rhythm regulation and would provide insights on the molecular-based adaptations of photoreception during vertebrate evolution. The goal of this study was to assess the selection patterns and evolutionary history of the melanopsin (*OPN4m* and *OPN4x*) paralogs at the gene and protein level. We tested the role of gene duplication and non-synonymous positively-selected substitutions in producing the extant diversity of physiological responses of melanopsin in both visual and non-visual photoreception organs and assessed the selective pressures on the retinal-related sites that determine the spectral absorption of melanopsins and the IL3 and IL2 that are involved in signalling light at the intracellular level. We also described the lineage-specific duplication that occurred in teleost fish that conferred novel photic capacities in new photic environments. Finally, we investigated the physiological plasticity of melanopsins by inferring the G-protein coupling proclivities of each gene.

Results

The Evolutionary History of Melanopsins

To understand the origin of melanopsin protein family, 51 *OPN4* gene sequences were retrieved from the Ensembl and NCBI databases from the main groups exhibiting melanopsins, including echinoderms and chordates (**Table S1**). The sequences were obtained by blasting both annotated-sequence databases and non-annotated genomes. To describe the emergence of melanopsin we compared available rhabdomeric photoreceptor sequences, including both melanopsin and invertebrate rhodopsin genes. Rhabdomeric photoreceptors comprehend two distinct evolutionary lineages: the *InRHO* that are present in protostomes, and the *OPN4* from deuterostomes [24]. Although our phylogenetic analyses support this partitioning, we found that the echinoderms comprise the basal branch for rhabdomeric photoreceptors. However, we cannot determine at this time whether it is a true member of the melanopsin gene family or perhaps another rhabdomeric photoreceptor type that has not yet been described. Moreover, rhabdomeric photoreceptors showed a considerable degree of amino acid variability (0.307 ± 0.027 in *InRHO*, 0.580 ± 0.014 in *OPN4x* and 0.614 ± 0.021 in *OPN4m*) relative to their ciliary relatives (0.175 ± 0.020 in *RHO*).

There were several amino acid patterns that broadly track opsin function and structure during rhabdomeric photoreceptor evolution (**figure 1**). Notably, echinoderms presented a FRY motif instead of the characteristic DRY motif of the rhabdomeric family, the E counterion found in all rhabdomeric opsins is replaced by an A in echinoderms and the stability residues of the CD8 domain had an analogous substitution in arthropods and vertebrates (F→Y). Furthermore, we inferred the maximum-likelihood ancestral sequence of the rhabdomeric ancestor and the most-likely ancestral characters of the DRY, Y and E counterions and the NP-Y-F motifs. Remarkably, these are the same amino acid motifs found in the rhabdomeric photoreceptors of extant annelids, mollusks and cephalochordates.

Despite the fact that we found melanopsin representatives in cephalochordates and vertebrates, BLAST searches of the available urochordate (*Ciona intestinalis* and *C. savignyi*) genomes, nucleotide collections and expression sequence-tag libraries were inconclusive (no sequence matches were retrieved with a high similarity level). The phylogenetic tree of vertebrate melanopsins (**figure 2A**) highlighted melanopsin evolutionary history, which included the duplication events leading to the origin of the *OPN4m* and *OPN4x* paralogs (2R, second round of whole genome duplication) and the teleost fish duplications leading to *OPN4m1*, *OPN4m2*, *OPN4m3*, *OPN4x1* and *OPN4x2* (3R, third round of whole genome duplication) [8,13]. These nodes are supported by high bootstrap and posterior probability values (higher than 95 and 0.95, respectively).

Although we did not find a complete sequence of either *OPN4m* or *OPN4x* in the lamprey (*Petromyzon marinus*), our blast searches identified an incomplete DNA fragment (ENSPMAG00000006406) that resembled an *OPN4m* melanopsin variant and phylogenetic analyses grouped the sequence with the *OPN4m* clade with 94% bootstrap and a posterior probability of 1.00 (**figure S1**). Since lampreys are one of the basal groups of vertebrates, this suggested that the melanopsin duplication event occurred earlier, before the emergence of cyclostomes. Also, our synteny analyses showed that the lamprey OPN4 genomic neighborhood includes the *LDB3* gene, which is congruent with observed patterns in the *m*-type paralog found in all other vertebrate taxa (**figure 2B**).

In the monotreme platypus (*Ornithorhynchus anatinus*) genome, our blast searches only found evidence of the *OPN4m* (ENSOANG00000010446) variant, indicating that the *OPN4x* variant was lost early in mammalian evolution, corroborating previous findings that suggested the absence of the gene in the marsupial *Sminthopsis crassicaudata* and placental mammals [11]. Therefore, the *PGDS-OPN4x-PDLIM5* paralogon found in all tetrapoda, could be different in mammals because *OPN4x* was lost earlier in the mammalian ancestor (**figure 2B**). This hypothesis is supported by: (i) our synteny analyses that showed the absence of the *OPN4x* gene in the genomic segment between the *PGDS* and *PDLIM5* genes in all mammals (monotremes, methatheria and eutheria), (ii) by our blast searches in mammals that did not retrieve any matches with the OPN4x protein and (iii) because the *OPN4x* transcript was missing in the very exhaustive human and rat expressing sequence tags databases.

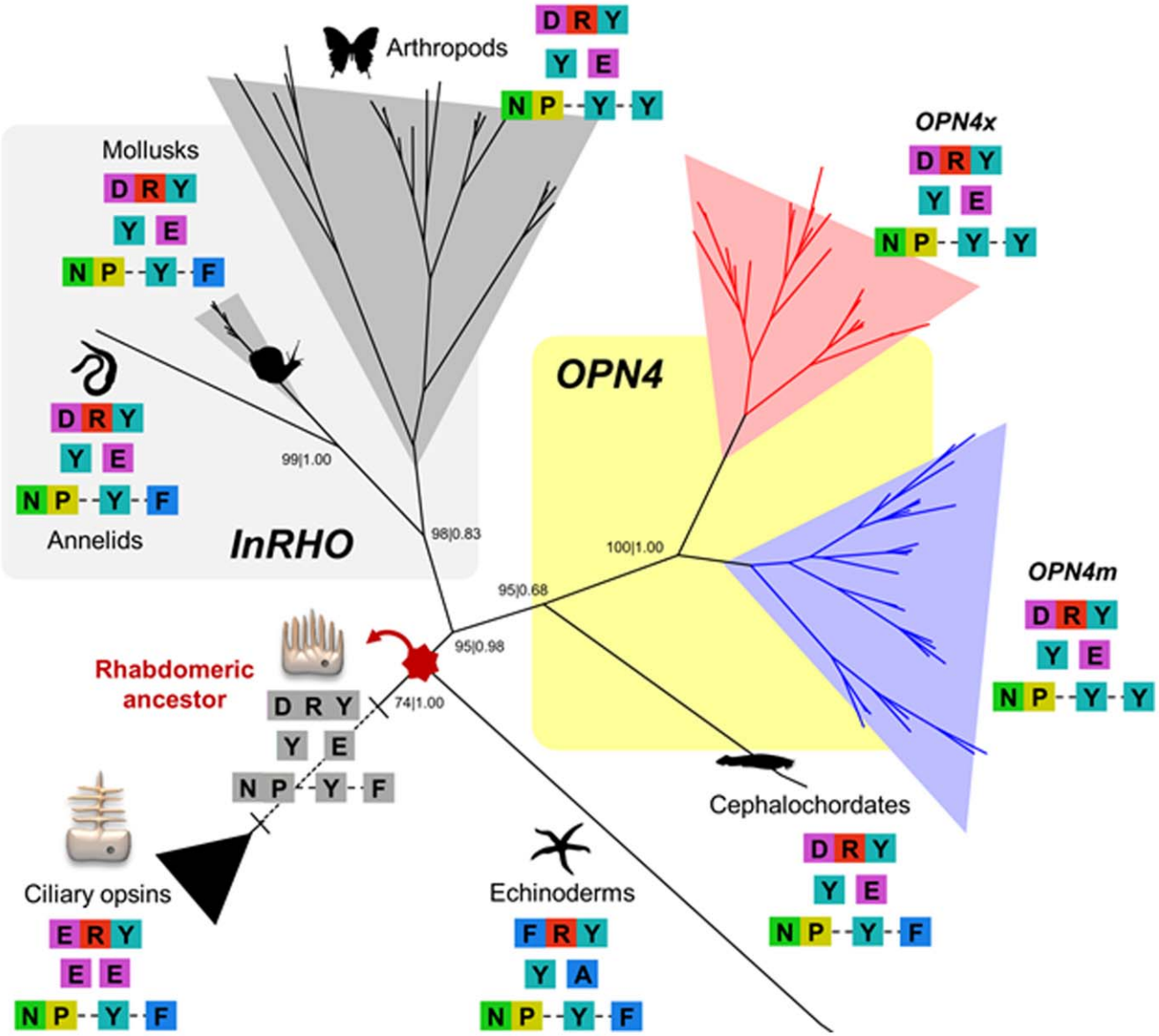


Figure 1. Phylogenetic depiction of the common-ancestry of invertebrate rhodopsins (*InRHO*) and melanopsin. The main opsin amino acid substitutions which are critical for the protein functional and structural innovations are color-coded. Maximum likelihood (ML) and Bayesian methods were used to build the phylogenetic tree and the support values of each method are shown for the main nodes (bootstrap and posterior probability, respectively). The grey amino acids are the maximum likelihood predicted motifs of the rhabdomeric photoreceptor ancestor.
doi:10.1371/journal.pone.0052413.g001

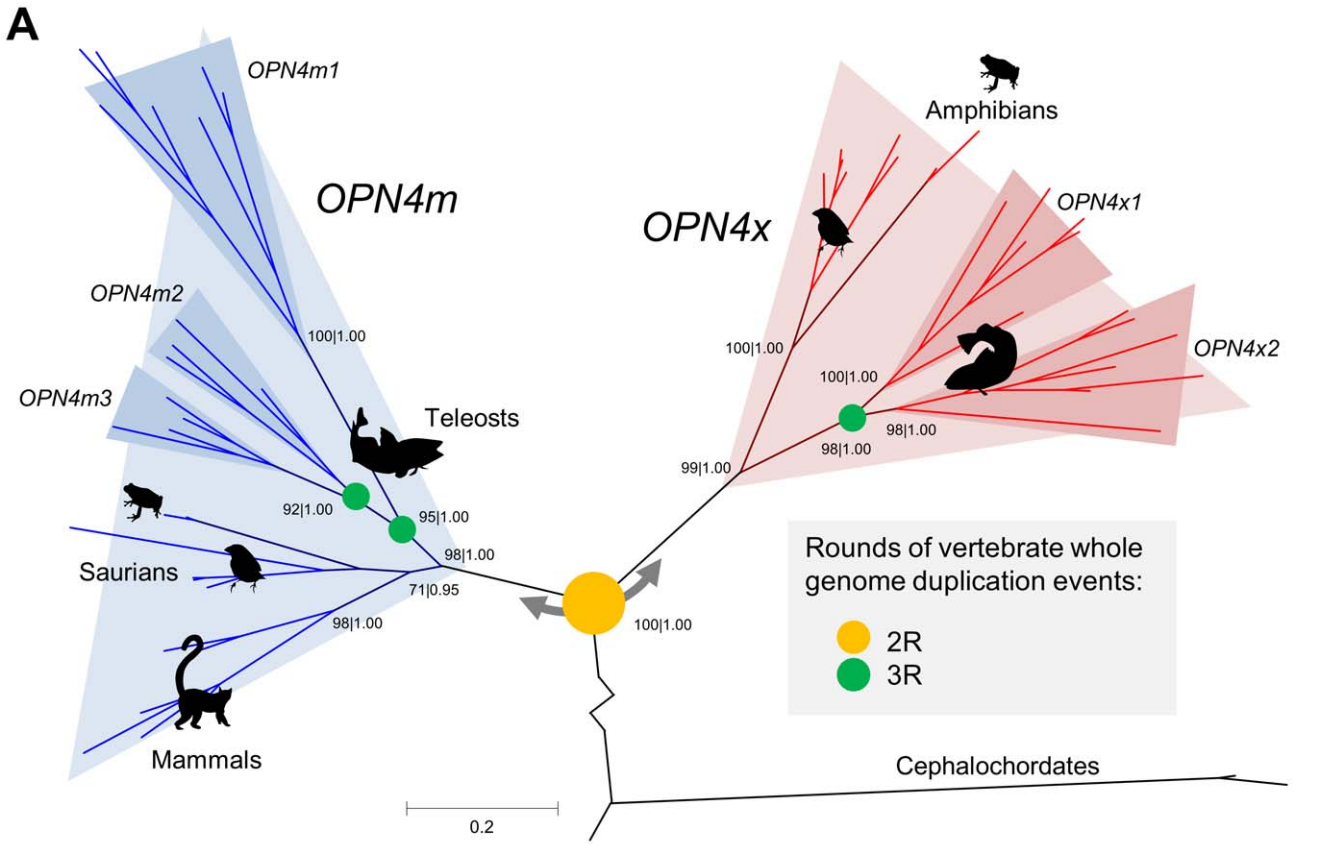
The Onho hypothesis advocates that two rounds of whole genome duplication occurred between the origin of chordates and the origin of jawed vertebrates, likely explaining the great number of paralogous genes in vertebrate genomes [25]. The existence of the *OPN4m* and *OPN4x* paralogs in vertebrate genomes, in addition to our evidence of the *m*-paralog in the lamprey genome, is consistent with the 2R event (**figure 2B**).

Whole genome duplication events shaping the genomes of vertebrates have not only been proposed in the early evolution of vertebrates, but also in the stem lineage of teleost fish, after their divergence from the land vertebrates (3R) [26]. We advanced that the melanopsin lineage specific duplications found in teleost fish (*OPN4m1*, *OPN4m2*, *OPN4m3*, *OPN4x1* and *OPN4x2*) probably occurred around 320 mya (3R event, **figure 2B**) [27,28].

Selective Pressure and Conservation in Melanopsins

Evidence of positive or negative selection at specific amino acid residues in vertebrate melanopsins was assessed based on the ratio of nonsynonymous (*d*N) versus synonymous (*d*S) substitutions (*d*N/*d*S or ω). A ω value less than 1 is indicative of purifying selection acting against amino acid changes, whereas a ω value greater than 1 suggests an excess of amino acid changes, indicative of adaptive evolution [29]. To test for positive selection at individual nucleic acid codons we used the site-specific models implemented in codeml program of PAML v4 package [30].

There was no evidence of significant positive selection at the nucleotide site level in *OPN4m* or *OPN4x* under model M8 of positive selection. Similarly, the global ω value under model M7 of no positive selection was very low in both cases (0.172 in *OPN4m* and 0.170 in *OPN4x*, **table 1**) indicating that the evolution of


A
OPN4m1
OPN4m
OPN4m2
OPN4m3
100|1.00
Teleosts
92|1.00
95|1.00
98|1.00
Saurians
71|0.95
98|1.00
Mammals
100|1.00
OPN4x
Amphibians
OPN4x1
OPN4x2
100|1.00
100|1.00
98|1.00
98|1.00
99|1.00
Rounds of vertebrate whole genome duplication events:
2R
3R
Cephalochordates
0.2

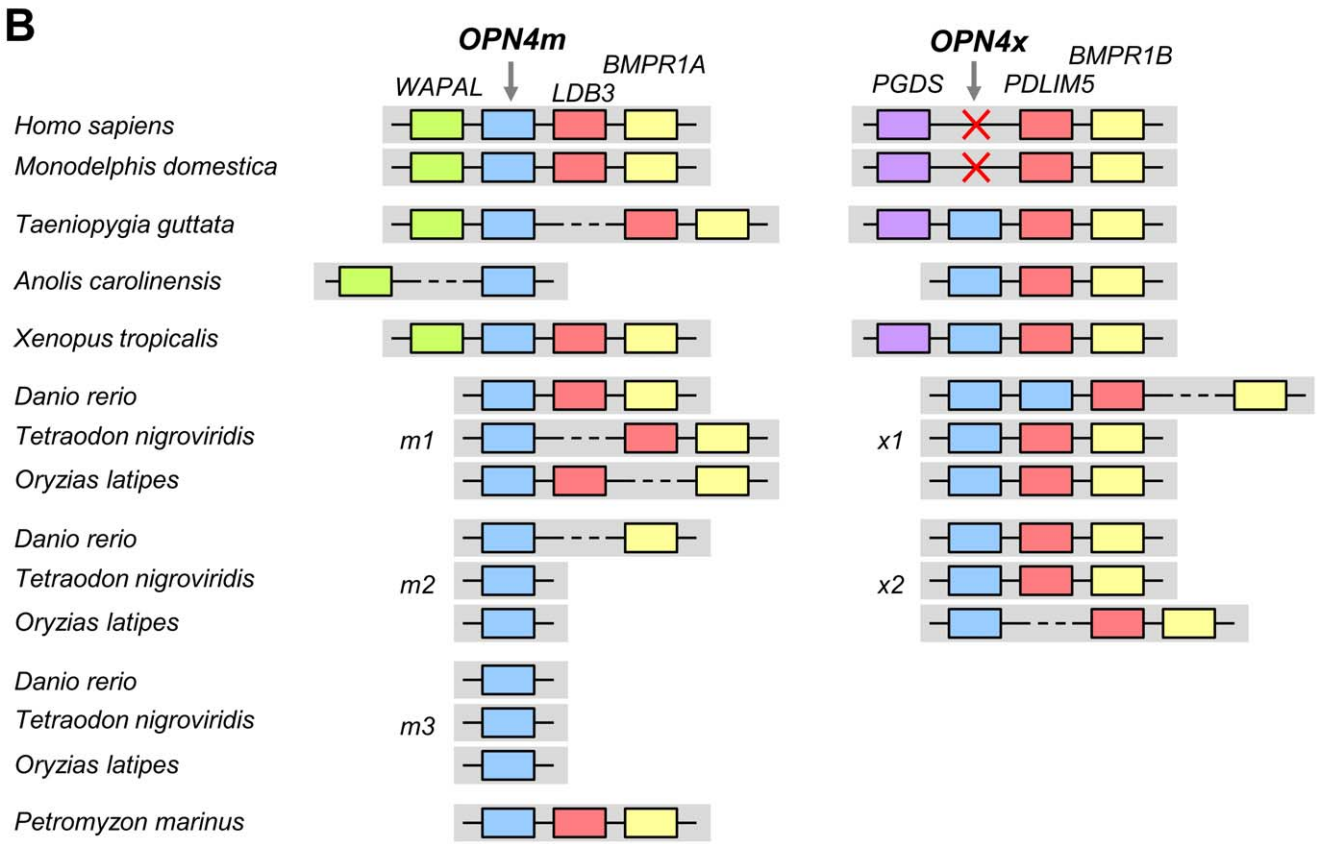

B
OPN4m
WAPAL
LDB3
BMPR1A
OPN4x
PGDS
PDLIM5
BMPR1B
Homo sapiens
Monodelphis domestica
Taeniopygia guttata
Anolis carolinensis
Xenopus tropicalis
Danio rerio
Tetraodon nigroviridis
Oryzias latipes
m1
x1
Danio rerio
Tetraodon nigroviridis
Oryzias latipes
m2
x2
Danio rerio
Tetraodon nigroviridis
Oryzias latipes
m3
Petromyzon marinus

Figure 2. Melanopsin gene tree and the syntenic analyses in the melanopsin genomic paralogon. A. The phylogenetic analyses were retrieved with maximum likelihood and Bayesian methods and the support values for each method (bootstrap and posterior probability, respectively) are shown on the main nodes. The main duplication events that characterize melanopsin gene history are represented with yellow (2R) or green (3R) circles on the respective nodes. B. Paralogous genes are represented with the same color code (*LDB3/PDLIM5* and *BMPR1A/BMPR1B*). The red cross represents the gene loss in the mammalian *OPN4x*.
doi:10.1371/journal.pone.0052413.g002

melanopsins in vertebrates was constrained by very stringent selective pressure. Our results contrasted the analyses performed by Dong *et al.* (2010), which reported a 0.07 global omega value for melanopsins [31], largely because we performed separate tests for each melanopsin paralog, used fewer mammalian sequences to reduce saturation bias in our alignments and because we implemented the more-appropriate M7–M8 test-comparison to infer negative selection instead of M8a-M8. The neutral M8a model implements an omega value that is fixed and equal to 1 [32], allowing the discrimination between neutral or positive selection.

To further assess selective pressure among sites and to characterize the slow- or fast-evolving domains of melanopsins, we plotted the variation of the ω value for the *OPN4m* and *OPN4x* codon-sites (**figure 3A**). This demonstrated that despite the strong purifying selection experienced by the *OPN4x* and *OPN4m* paralogs, some regions of the molecules accumulated non-synonymous variation. To avoid overestimating the ω value on the N terminus, since some sites are not fully represented for all taxa, we excluded the final part of the N terminus on the **figure 3A** diagrams.

The Mann-Whitney test was used to calculate W statistics [33] and to test the alternative hypothesis of significantly greater median-ω-values in the suspected regions. We tested the ranks of the suspected sites (n) against the remaining sites (N – n) using the same total number of sites for each paralog (N = 420). The C and N-terminus melanopsin domains evolved at higher ω-values (W = 30304*, n = 123 in *OPN4m* and W = 33375*, n = 126 in *OPN4x*), suggesting more amino acid variability in terminal regions. Also, a higher ω value was observed in the second and the third intracellular loops (IL2 and IL3) as well as the helix bundles that comprise each loop (W = 13541.5*, n = 58 in *OPN4m* and W = 13109.5*, n = 63 in *OPN4x*). Together, these regions (plus the CD8 domain) interact with the G-protein that mediates the phototransduction pathway [34].

Table 1. Site-specific selection models for the vertebrate melanopsin *OPN4m* and *OPN4x* genes.

Gene	Model	ω	*lnL*	Hyphothesis	LRT	*df*	
OPN4m	A. M0	.132	−17125.136				
	B. M3	.170	−16379.268	A *vs.* B	1491.736	4	*
	C. M7	.172	−16377.538				
	D. M8	.172	−16377.500	C *vs.* D	0.077	2	
OPN4x	A. M0	.127	−13962.083				
	B. M3	.168	−13413.817	A *vs.* B	1102.532	4	*
	C. M7	.170	−13408.551				
	D. M8	.177	−13406.677	C *vs.* D	3.748	2	

The likelihood values and the respective estimated parameters are shown for each model. The ω ratio is an average over all sites of the *OPN4m* and *OPN4x* paralogs. The asterisk (*) means that the alternative hypothesis is statistically significant at a 5% level, implementing the LRT (likelihood ratio test). Notes: *df* – degrees of freedom.
doi:10.1371/journal.pone.0052413.t001

Further insights on the relationship between melanopsin structure and function were obtained through a protein-level approach by combining information from the three-dimensional melanopsin structure and the physico-chemical properties of the amino acid substitutions. TreeSAAP v3.2 was used to reconstruct ancestral sequences and to determine and categorize evolutionary changes in 30 amino acid properties [35]. We looked for positively selected sites under destabilizing selection (non-synonymous substitution with significant disequilibrium changes to the molecule) and found that 70% of the substitutions had probable chemical implications and 30% had structural implications in both paralogs (**Table S2 and Figure S2**). As expected, substitutions that potentially changed chemical properties were more common than substitutions with structural implications. Thus the heptahelical conformation of melanopsins was safeguarded throughout evolution.

27 and 21 sites were under destabilizing positive selection in both OPN4m and OPN4x, respectively (**figures 3B and 3C**). A chi-square adjustment test with a 5% level cutoff showed that destabilizing positive selected sites had a differential distribution between the extra and intra-membrane regions of the protein ($\chi^2 = 10.703$* in OPN4m and $\chi^2 = 5.762$* in OPN4x, both tested at 1 degree of freedom). A large proportion of sites under destabilizing positive selection were located in the IL2 and IL3 and in the helix bundles that comprise each loop (**figures 3B and 3C**). This pattern is more evident in OPN4m (15/27 = 0.56) than in OPN4x (5/21 = 0.24). The predicted three-dimensional conformation of melanopsin showed that these specific sites are located on the intracellular part of the molecule where the G-protein interaction is established. As in the results obtained in the site selection analysis, the conservation index estimated on the Consurf webserver [36] showed that (i) both the N and C terminus are highly variable, (ii) the second and third intracellular loops are unexpectedly variable and (iii) the molecule interior, responsible for the retinal accommodation, is very conserved (see detailed aspects in **figures 3B and 3C**). The proportion of variable sites on the melanopsin molecule was around 55% in *OPN4m* and 59% in *OPN4x*.

OPN4 Duplications and Functional Divergence

Melanopsin evolutionary history has been marked by a series of gene duplications episodes (**figure 2**). Therefore, we tested for branch and branch-site selection for the main duplication events of melanopsins (*OPN4m/OPN4x*, *OPN4m3/OPN4ma* and *OPN4x1/OPN4x2*). In addition, we assessed the type I and type II functional divergence between variants using Diverge v2.0 [37]. Type I functional divergences represent amino acid configurations that are highly conserved in one clade, but are variable in the other clade, denoting residues that have experienced differentiated functional constraints at a particular site. Type II represent residues which are very different between clades, but are found in very conserved amino acid configurations in both clades, implying that these residues may be responsible for functional specification, especially when the substitution has some biochemical significance [38]. Type I and type II functional divergence tests for each group of duplicates are summarized in **table 2** and the additional information on the branch and branch-sites tests, the estimated

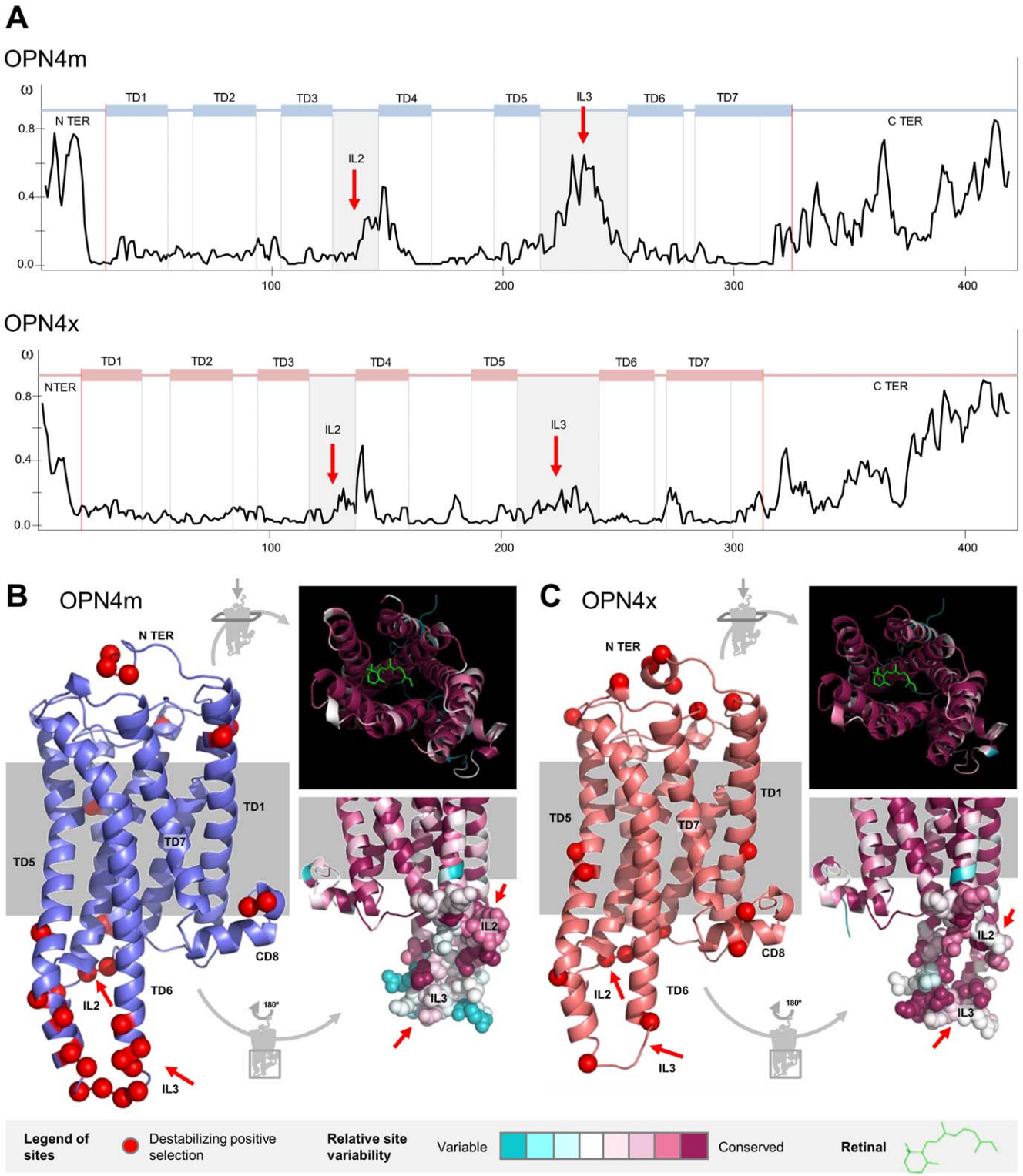


Figure 3. Destabilizing positively selected sites and conservation index in the *OPN4m* and *OPN4x* paralogs. A. ω-ratio site estimation for each melanopsin paralog. The IL2 and IL3 regions are highlighted (red arrows). **B and C.** Three-dimensional structure of *OPN4m* and *OPN4x* paralogs showing the sites under positive destabilizing selection (red) and detailed perspectives of the conservation index in the interior of the molecule, where the retinal is accommodated, and the IL2 and IL3 loops (red arrows).
doi:10.1371/journal.pone.0052413.g003

parameters and the inferred selected amino acid sites are presented in **table S3**. All the numerical and amino acid identification of sites are based on the *Gallus gallus* OPN4m and OPN4x protein sequences.

Table 2. Type I and type II divergence between the OPN4 paralogs and the teleost lineage-specific duplications.

	OPN4x/*OPN4m*	*OPN4x2*/*OPN4x1*	*OPN4m3*/*OPN4ma*
Residues	294	339	330
$\theta_I \pm se$	0.349±0.059*	0.039±0.082	0.181±0.082*
z_I	6.362	0.712	3.336
p–value	0.000	0.238	0.000
$\theta_{II} \pm se$	0.126±0.084*	0.044±0.058	0.048±0.061
z_{II}	2.166	0.799	0.874
p–value	0.016	0.212	0.191

θ_I and θ_{II} are the coefficients of type I and II functional divergence. Asterisks (*) mark results with statistical significance at 5% level of confidence and *se* denotes the standard error.

doi:10.1371/journal.pone.0052413.t002

After the *OPN4x*/*OPN4m* duplication event, the number of non-synonymous substitutions increased which led to a higher overall ω-ratio on these lineages. 25% of the melanopsin sites were under positive selection in the *OPN4x* lineage. There was a significant functional divergence between *m*-melanopsin and *x*-melanopsin, indicated by 6% and 8% of the sites being under type I and type II functional divergence, respectively. Positively selected sites, as those involved in type I and type II functional divergence on the *G. gallus OPN4m* three-dimensional structure are displayed graphically in **Figure 4**. A group of residues on the initial regions of the TD5 and TD4 (200F, 273S and 276A) that are involved in retinal connection showed evidence of functional divergence and/or positive selection (**figure 4C**). The IL3 and IL2 and the respective bundles both had sites with signals of positive selection or that contributed to functional divergence (e.g. 137A, 141V, 224K, 227K, 240E and 247R).

At least two whole duplication events were fundamental in determining actual teleost *m*-type melanopsin patterns (*OPN4m3*/[*OPN4m2*+*OPN4m1*], *OPN4m1*/*OPN4m2*), but only one even is sufficient to explain *x*-type evolution (*OPN4x1*/*OPN4x2*). To simplify the clade notation, when refer to *OPN4m2*+ *OPN4m1* clade instead as *OPN4ma*. Taking into account both phylogenetic and synteny analyses in teleost fish, we studied the three duplication events (**figure 2A**) of teleost melanopsin paralogs in more detail. Due to an insufficient amount of available sequences for the *OPN4m1* and *OPN4m2* duplicates, we have not done a branch-site or functional-divergence analysis for the *OPN4m1*/*OPN4m2* duplication event.

In the *OPN4m3* lineage 11% of the residues were under positive selection and both copies showed evidence of type I but not type II functional divergence. The main residues responsible for positive selection and functional divergence are located in the TD5 and the CD8 regions (**figure 5**). Moreover, we found that *OPN4m3* protein sequences of the DRY were replaced by the DRC motif. Both lineages of the *OPN4x1*/*OPN4x2* duplication were under positive selection, although to a lesser extent (around 5%), and no evidence of functional divergence was found between these copies.

G-protein Couple Receptors

Melanopsins process light by using a G-protein that establishes a physical-chemical interaction with the intracellular domains of the opsin. We used Pred-Couple v2.0 web server to determine the potential G-protein couple preferences of GPCRs on the four possible subfamilies (Gs, Gi/o, Gq/11 and G12/13) [39]. We found that melanopsins have a possible promiscuous interaction with two G-proteins: Gi/o and Gq/11. There was no evidence that G12/13 was a coupling G-protein, which increased confidence in the accuracy of our results, as this is a ciliary-type G-protein.

For the teleost fish melanopsin duplications, the *OPN4x1* copy showed affinity with the Gq/11-type and *OPN4m3* with the Gi/o, both with >0.90 posterior probability level (**figure 5**). Both x-type and m-type melanopsins in birds had affinity with the Gq/11 G-protein (0.89 and 0.84 in *OPN4m* and *OPN4x* on *Gallus gallus* amino acid sequences). In mammals, higher affinity was also observed for the Gq/11-type G-protein with a posterior probability of 0.96 and 0.91 in *Canis familiaris* (Laurasiatheria representative) and *Loxodonta africana* (Afrotheria representative), respectively. Therefore, Gq/11 was the most likely G-protein intervenient in the melanopsin phototransduction cascade, especially in non-fish vertebrates.

Discussion

Understanding the molecular evolution of photoreceptor genes is crucial to assessing how genetic variation influences molecular specialization and to understanding the implications to how organisms have adapted to different photic environments. At the molecular level melanopsins may have specialized by (i) establishing distinct coupling preferences with the signalling cascade in the cell interior and/or (ii) changing their spectral sensibility accordingly to environmental conditions. The implications of which are discussed below.

Integration of Light by Melanopsin – the Variability of the Second and Third Intracellular Loops (IL2 and IL3) and G-protein Type Preferences

Our evolutionary analyses of the rhabdomeric photoreceptors suggest an urbilaterian common-ancestor for both *OPN4* and *InRHO* orthologs (**figure 1**). This result corroborates the general Arendt theory of photoreceptor cell-type evolution [24,40] that supposes a rhabdomeric-like cell in the set of photoreceptors of the ancient urbilaterian eye. Additionally, the inferred ancestral amino acid sequence for the urbilaterian rhabdomeric ancestral photoreceptors suggests that the molecular basis of rhabdomeric-like light transduction remained similar to that observed now. Therefore, some extant groups (annelids, mollusks and cephalochordates) have the same combination of amino acid motifs (**figure 1**). This result supports the idea of a universal method of signalling light in the rhabdomeric photoreceptors, at least in the mechanisms of retinal biding and structural maintenance that these amino acid motifs perform.

Furthermore, experimental studies show that all rhabdomeric photoreceptors share the same signal transduction pathway, including the activation of the phospholipase C (PLC) and the inositol phosphate (IP3), which involves the Gq/11 G-protein type [17,18,20,41]. However, we determined that there is possible uncertainty in the affinity of teleost fish melanopsins relative to their G-protein couple preferences: Gi/o and Gq/11 (**figure 5**). It should be stressed that for the mammals and birds studied here, the Gq/11 was always predicted to be the most-likely intervening G-protein type. We propose that these promiscuous coupling preferences in teleost fish may constitute an evolutionary advantage since one environmental signal may produce a great quantity of internal organism responses. We suggest that this behavior may provide an ecological advantage by originating new and more complex photo-irritability responses to environmental stimuli. Moreover, we observed unexpected variability in the IL2 and IL3 loops suggesting, in agreement with the previously-discussed result, the ambiguous activation of more than one G-

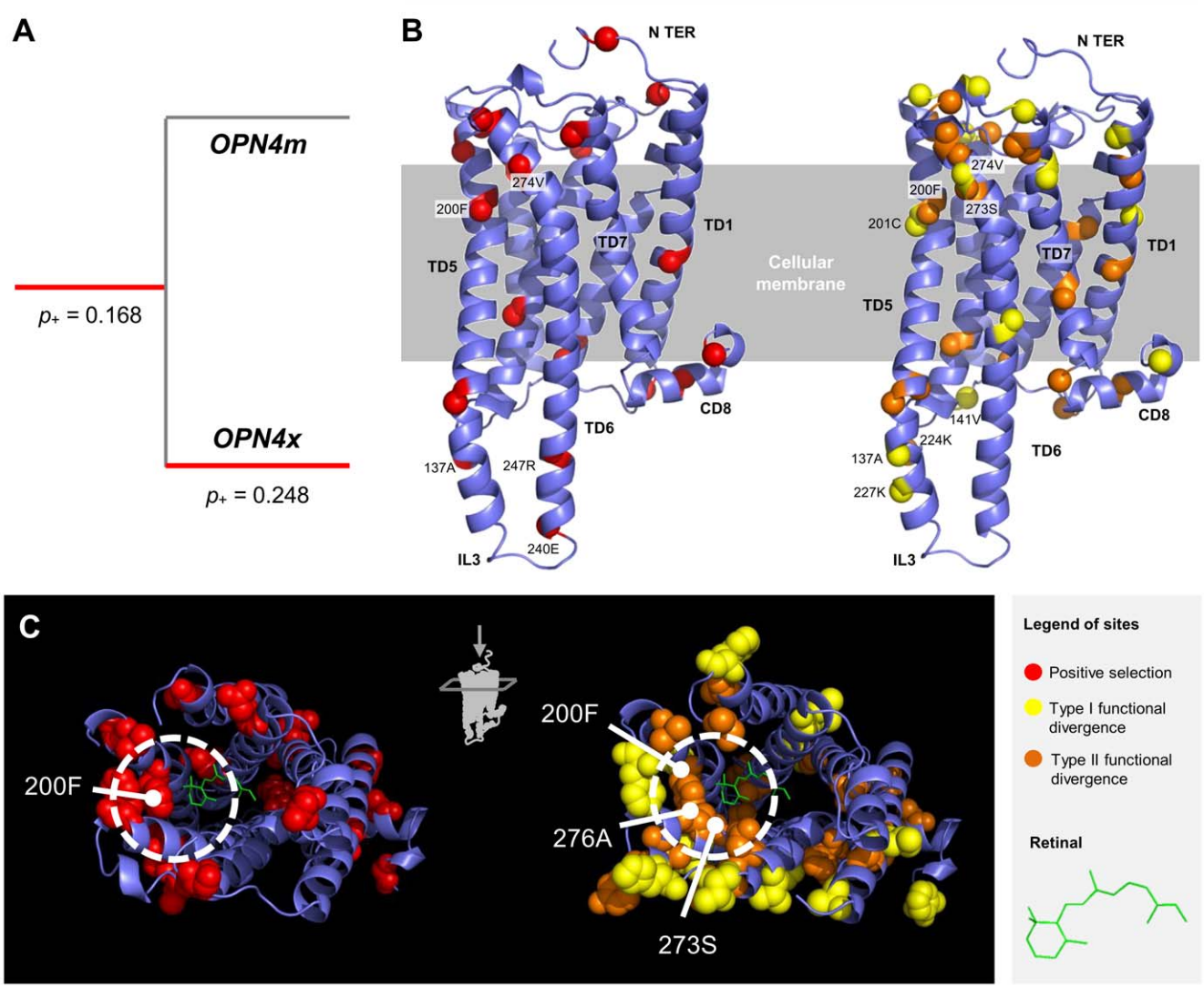


Figure 4. Branch and site selective pressures during the *OPN4m/OPN4x* duplication event. A. Branch-site tests. Red lineages represent an inferred episode of positive selection. In those branches is represented the *p*+ parameter (proportion of the positively selected sites). **B.** Representation of the positively selected and functional divergence sites (type I in yellow and type II in orange) in the three-dimensional structure of the *Gallus gallus* OPN4m protein. **C.** A detailed perspective of the retinal accommodation on the melanopsin molecule and the occurrence of the positively selected and type I and II functional divergence sites.
doi:10.1371/journal.pone.0052413.g004

protein. We advance three possible resolutions to this quandary: (i) Gq/11-type G-proteins do not require conserved intracellular domains to establish a coupling ligation in melanopsins, (ii) intracellular loop variability contributes to G-protein coupling promiscuity on melanopsins, or, a less-likely but possible explanation that (iii) another type of G-protein mediates the melanopsin phototransduction pathway.

OPN4m and *OPN4x* Paralogs and the Emergence of the Complex Eye in Vertebrates

We found that melanopsins were apparently lost in tunicates, whereas only one copy is present in cephalochordates and vertebrates present two copies. Gene loss in urochordates is generally assumed to be common, and it was already reported for the well-studied *Hox* genes [42,43] so we hypothesize that melanopsin may have been lost during a genomic rearrangement process. However, regardless of the quality of the genome assembly, it should be noted that negative results from gene searches in genomes or DNA libraries may be biased because of incomplete genome sequence, the lack of protein homology or missing sequence data. To date, the *Ci-opsin1* and the *Ci-opsin2* ciliary opsin genes involved in photic stimuli in larval stages have been identified in *C. intestinalis*, but other types of photoreceptors cells have also been identified [44,45]. More molecular studies are needed to more-thoroughly evaluate the presence or absence of a rhabdomeric-like photoreceptor in urochordates genomes, which would be of great importance in disentangling the ancestral photoreceptor content of the vertebrate eye.

All vertebrates have anatomical features that are not observed in their closest living relatives, the urochordates and cephalochordates. It has been shown that the 1R and 2R whole genome duplication events seem to explain the photomorphological diversity that we can currently see in vertebrates [46]. Cyclostomes are a very basal group in vertebrate phylogeny and the presence of

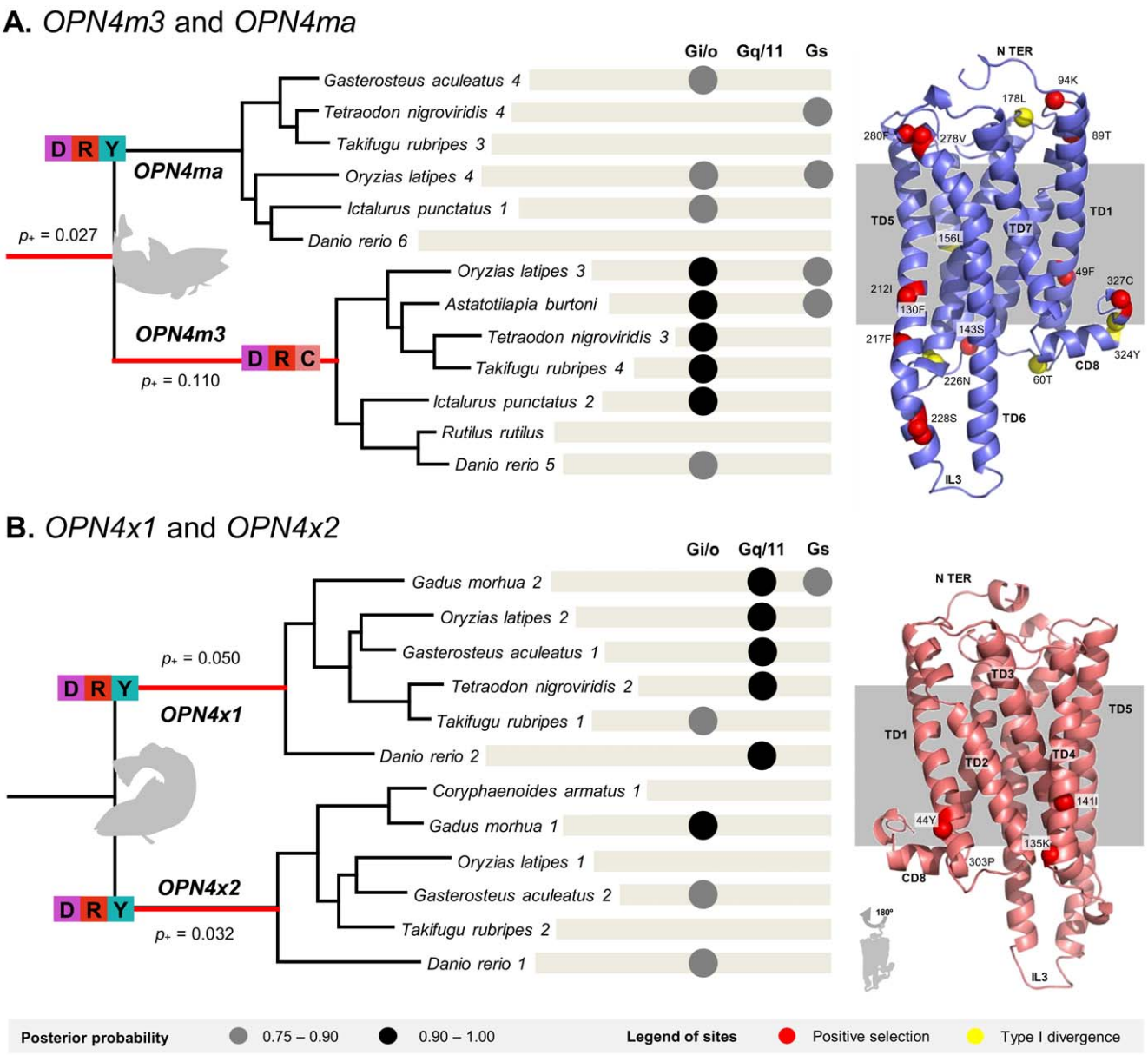


Figure 5. Branch and site selective pressures during the teleost lineage-specific duplications: A. *OPN4m3/OPN4ma* and B. *OPN4x1/OPN4x2*. A punctual substitution (Y→C) was determined in the DRY motif in the OPN4m3 teleost melanopsin duplicant. Red lineages represent an episode of positive selection and the *p*+ parameter means the proportion of the positively selected sites. Black and grey circles represent the posterior probability level of G-protein coupling preference for each teleost fish amino acid sequence: 0.75–0.90 (grey circles) and >0.90 (black circles). The three-dimensional structure of the *Gallus gallus* OPN4m and OPN4x paralogs is also represented showing the occurrence of positive selection and functional divergence at the site level.
doi:10.1371/journal.pone.0052413.g005

the *OPN4m* variant introduced by us (**figure S1**) is consistent with a whole genome duplication event just before the emergence of jawless fish, coincident with the 2R episode. Moreover, the presence of genomic paralogons among vision-related genes produced by the 2R episode seems to be common pattern in visual opsins, as has been demonstrated through the study of the protein intervenes in the vertebrate visual cascade [47]. The syntenic and phylogenetic analyses of *OPN4m* and *OPN4x* (**figure 2**) suggest that a whole genome duplication event occurred during the emergence of vertebrates, as with the 2R episode. These result predicts that the emergence of melanopsin variants parallel the vertebrate emergence (at least 600 mya), earlier than the origin of the Tetrapoda in the Late Devonian (360 mya) as proposed by Bellingham et al. (2006) [8].

However, the question remains as to why both paralogs were maintained in the genome following the duplication event. We hypothesize that an advantageous dosage effect can explain the retention of the duplicated melanopsin paralogs in the genome [48,49]. We assume that a photoreceptor dosage effect could have been be of great advantage, or at least more advantageous than the expected metabolic constraints such as energy loss and the regulation of the signalization pathways. Not only the organization of the non-visual system went through dramatic changes during the emergence of vertebrates, but the visual system also changed

significantly, as demonstrated by the photoreceptors and their current paralogs (e.g. rhodopsins and conopsins), and as such are arguably the principle reason for the development of complex eye novelty [50–52]. Thus, the complex visual system of vertebrates is the result of the large number of photoreceptors that enable the processing of wavelengths in different ranges of the light spectra (visible and also UV). The melanopsin group, as well as the ciliary opsins (e.g. rhodopsins and conopsins), show diverse duplicated copies (*Rh1*, *Rh2*, *SWS1*, *SWS2* and *LWS*) that over time underwent further specialization, and which presently regulate important processes such as color-vision or circadian-rhythm synchronization [53].

Melanopsin and Site Level Selective Pressures – Evidence of Spectral Sensibility Specialization

Our results show that melanopsin amino acid substitutions are mainly under negative selection. This suggests that melanopsins play an important physiological role in the photoreception system and that the complete or partial loss of melanopsin functionality would compromise organism fitness. Indeed, mammalian melanopsin is responsible for phase-shifting circadian rhythms, plasma melatonin suppression, spanning pupil constriction and the dependent irradiance regulation of retinal cone function [54]. These functions are related with basic physiological needs, such as feeding and reproduction, thus justifying the need of fine-scaled regulation at the genetic level. Among non-mammalian vertebrates, several photoreceptive locations have been well-described in addition to the retina, including the pineal gland and deep brain [55,56]. In these extra-retinal photoreceptors, the role of melanopsin is not completely understood, and since both melanopsin paralogs are present in non-mammalian vertebrates, inferences of selective pressure acting in these lineages should be made with caution. Despite the indication of a general purifying selective signature mediating melanopsin evolution, we identified several sites that are responsible for both selective and functional divergence between the *m* and *x* melanopsins. The *OPN4x* lineage showed evidence of positive selection, which suggests a relaxation of the selective pressure favoring genetic variation following the post-duplication episode. Additionally, functional divergence types I and II were detected, indicating a process of functional differentiation and specialization over 600 mya of vertebrate evolution. Indeed, we show that some sites under positive selection and functional divergence near the retinal localization (200F, 273S and 276A) (**figure 4**) with likely implications to spectral sensibility. It has been shown that in cones and rhodopsins the sites responsible for spectral tuning tend to cluster around either the Schift base linkage or the ionone ring of retinal [57]. In contrast, in chicken (*G. gallus*) the m and x-type melanopsins showed the same spectral sensibility (476–484 nm) [58]. However, zebra fish spectral sensitivity for *OPN4m3* and *OPN4x2* is highest at 484 *nm* and 470 *nm*, respectively [13].

The 3R Event and the Large Number of Melanopsin Paralogs in the Teleost Eye

We hypothesize that melanopsin copies may have been key to the radiation of teleost fish (3R event, **figure 2B**), playing a major role by providing new photic capacities in new environments. Aquatic environments are very complex from the photic point of view, varying based on numerous factors including turbidity, salinity, pressure and depth that result in very different refractive indexes throughout the water column [59]. Thus, the existence of many photoreceptors would be an advantage in such complex ecosystems. Interestingly, we identified five melanopsin representatives in the teleost retina while most of vertebrates have two, implying the existence of a complex non-visual signalling pathway in teleost fish and the involvement of multiple protein complexes.

Moreover, it is known that *OPN4m3*, *OPN4x1* and *OPN4x2* are monostable photopigments, while instead *OPN4m1* and *OPN4m2* display invertebrate-like bistability [13]. Bistable pigments are thermally stable before and after photo-activation, but monostable pigments are stable only before activation [60]. Accordingly, our results suggest that a process of functional divergence and diversifying positive selection occurred on the *OPN4ma* (*OPN4m1*+ *OPN4m2*) and that *OPN4m3* is located mostly on the TD5 and CD8 domains (130F, 156L, 178L and 324Y) (**figure 5**). These domains may play an important role conferring structural ability to these pigments to perform the monostability or bistatibilty types of retinal accommodation. Indeed, CD8 domain is known to be involved in conferring structural integrity upon pigment activation [61]. Furthermore, OPN4m3 protein presents a substitution on the DRC motif, which may have implications to provide the negative charge to stabilize the inactive opsin. For the *x*-type duplications, we did not find any type of functional divergence between *OPN4x1* and *OPN4x2*, which are both monostable photopigments.

Conclusions

Our general results suggest that the main phenomena determining melanopsin gene family evolution are (1) purifying negative selection and (2) the duplication events followed by minor episodes of positive selection and functional divergence.Negative selective pressures help maintain the structural and biochemical homology observed among all opsin photoreceptors and duplication events are the source of gene number variation in the vertebrate genomes. In addition, the variability at the amino acid level is mostly located at the retinal biding-related sites and in the third and second intracellular loops. This suggests that vertebrate melanopsin adapted to new photic environments by one or both of these processes: providing sensibility to different quality and quantity of light and/or supplying new or more complex photo-irritability responses.

Methods

Data Collection

PSI-BLAST and TBLASTN searches with protein sequences of the two *Gallus gallus* melanopsins (NM_001044653.1 and AB255031.1) were performed in the NCBI data base [62] and the Ensemble genome projects [63]. 54 previously published sequences were collected, representing 26 different species from the main phylogenetic groups of the chordates phylum: two cephalochordates, 10 fish, two amphibians, six reptiles and birds and six mammals. **Table S1** shows the species names and reference numbers for each collected sequence. Two melanopsin sequences from the sea urchin (*Strongylocentrotus sp.*) were included as outgroups. All the sequences from the *InRHO* photoreceptors were retrieved from the Davies *et al.* 2010 [16].

Sequence Alignments and Phylogenetic Trees

A protein-based coding-sequence alignment was performed with the translated nucleotides sequences and the standard options of the Muscle version-3.3 algorithm [64], which was subsequently improved by manual inspection of the alignment. The quality of the alignment was enhanced with the Gblocks web server [65] by removing ambiguous and gaps-rich sites (>75% gaps). We then used three alignment sets in further analyses: (i) the default settings; (ii) eliminating sites with more than 75% of gaps; and (iii) removing

gap-rich sites but considering the codon information (used for the positive selection analyses).

The presence of saturation in base substitution for the *OPN4* and *OPN4m* and *OPN4x* variants was tested by comparing half of the theoretical saturation index expected when assuming full saturation ($I_{SS.C}$, critical value) with the observed saturation index (I_{SS}) [66]. No evidence of saturation in any of the referred alignments (**table S4**). jModelTest version 0.1.1 [67] implementing the Akaike Information criterion (AIC) was used to estimate the most appropriate model of nucleotide substitution for tree construction analysis. This procedure was repeated for each melanopsin paralogs genes, with the *OPN4m* and the *OPN4x* sequences. GTR+I+Γ was determined as the best-fit model for *OPN4*, *OPN4m* and *OPN4x* alignments. The estimated parameters under the selected nucleotide substitution model for each gene can be seen in **table S4**.

Phylogenetic trees were constructed using two distinct algorithms, Maximum likelihood (ML) in PhyML [68] and Bayesian analysis in Mr. Bayes 3.1.2 [69,70], using the estimated parameters found for the nucleotide evolutionary model determined earlier. Bootstrap analyses (1000 replicates) were used to assess the relative robustness of branches of the ML tree [71]. Bayesian analysis was conducted using the estimated parameters of the nucleotide substitution model as priors for 5.000.000 generations. Two concurrent runs were conducted to verify the results. The first 12500 trees were discarded as burn-in samples, the remaining trees were used to compute a majority-rule consensus tree with posterior probabilities. Synteny analyses were performed using the Ensembl and Genomicus version 64.1 data bases [63,72].

Positive Selection Assessment

OPN4, *OPN4m* and *OPN4x* alignments and the ML/Bayesian trees were used in the program codeml from the PAML version 4.4 software package [30] to assess the selective pressure acting on melanopsin sites. To examine the *dN/dS* or ω ratio, three codon substitution models of maximum likelihood analysis were performed: branch-specific, site-specific and branch-site likelihood models.

The site specific models were tested comparatively [73]: M0 (one ratio) *versus* M3 (discrete), M1a (nearly neutral) *vs* M2a (positive selection) and M7 (beta) *vs* M8 (beta+ω). Subsequent likelihood rate comparisons were performed to test which models fits the data significantly better. Model M0 assumed a constant ω-ratio, while in models M1a and M2a ω-ratio is supposed to be variable between sites. M7 and M8 assume a β-distribution for the ω value between 0 and 1. Models M2a, M3 and M8 allow the occurrence of positively selected sites. In addition, the ω value for each codon of the melanopsin *OPN4m* and *OPN4x* paralogs was assessed under the significantly selected site model, using the Selecton web server [74].

The branch selection models were implemented comparing the same ω ratio for all lineages in the tree (one-ratio model) and the two-ratio models assigned two ω ratios for the foreground (ω_1) and background branches (ω_0) [75]. The branch-site models allow the ω ratio to vary both among sites and among lineages and were used to detect positive selection that affects only a few sites along a few lineages. A most stringent branch-site test of branch-site test of positive selection was implemented comparing the alternative model A and the ω fixed null model [76]. When the likelihood ratio test was significant, the Bayes Empirical Bayes (BEB) method was used to calculate posterior probabilities of the sites that are subject to positive selection [77].

Branch-specific and branch-site models were implemented to study the melanopsin duplication event and both followed the approach outlined here: model A represents the selective pressure before the duplication event and models B and C had one ω value for each duplicated lineage following the duplication event. The significance for the referred likelihood ratio tests (LRTs) was calculated using the chi-square approximation $2\Delta lnL$, the double of the difference between the alternative and null model log likelihoods. LRT degrees of freedom are calculated as the difference of free parameters between the nested models.

A protein level analysis to detect possible positively selected sites were also investigated on the basis of 31 physicochemical criteria with TreeSAAP version 3.2 [35]. TreeSAAP measures the selective influences on structural and biochemical amino acid properties during cladogenesis, and performs goodness-of-fit and categorical statistical tests. The program classifies the range of changes in eight magnitude categories from conservative to radical for each amino acid properties and calculates a *z*-score that indicates the direction of selection (negative or positive selection) [78]. Positive radical or destabilizing selection sites (6, 7 and 8 magnitudes) as expected to result in significant structural and functional changes on the protein were monitored at the 0.01 significance level.

Structural Analysis and Homology Modeling

Three-dimensional homology models of melanopsin were built using Modeller version 9.9 [79] implementing a comparative protein structure by satisfying spatial restraints. Squid (*Todarodes pacificus*) rhodopsin protein data bank available structures 2ZIY [80] and 2Z73 [81] were selected as homology models. The predicted three-dimensional conformation of *Gallus gallus* m and x-type melanopsin was based on the invertebrate squid (*Todarodes pacificus*) rhodopsin protein 2ZIY [80]. Consurf webserver was implemented to calculate the conservation index and to assess the three-dimensional localization of most variable and conserved domains at the melanopsin molecule [36]. PyMol version 1.4 graphical interface was used to manipulate the melanopsin molecule and to perform all the images that include melanopsin three-dimensional structure [82].

Functional Divergence

Diverge version 2.2 was used to identify sites of type I and type II functional divergence, which occurs through changes in the amino acids biochemical properties at a specific positions between defined groups of related proteins [37]. The functional divergence between two monophyletic groups can be classified in two groups: (i) type I, if the amino acid pattern are very conserved in the duplicate gene but highly variable in the other gene copy, which implies shifted functional constrains and (ii) type II, when the amino acid pattern is very conserved in both the duplicated gene clusters but their biochemical properties are very different [83]. Type I and type II functional divergence was assessed by estimating the θ_I and θ_{II} divergent coefficients. θ parameter significantly greater than zero means that either altered selective constraints or a radical shift of amino acid physiochemical property after gene duplication is likely to have occurred [38,84]. A site-specific outline based on the posterior probability (>0.75) was used to predict critical amino acid residues that were responsible for functional divergence between groups. PredCouple 2.0 tool was implemented to predicted coupling specificity of GPCRs to the four known G-proteins families [39]. The predicted coupling specificity robustness of the melanopsin sequences was evaluated with the generated posterior probability.

Supporting Information

Figure S1 Melanopsin gene tree including the lamprey (*Petromyzon marinus*) blasted sequence ENSPMAG00000006406. ML and Bayesian method were performed to build the phylogenetic tree. Bootstrap and posterior probability support values are respectively represented for each node.
(PDF)

Figure S2 Comparative importance of destabilizing positive selected substitutions in the *OPN4m* and *OPN4x* paralogs for each amino acid property.
(PDF)

Table S1 Melanopsin sequences used in the phylogenetic analysis.
(PDF)

Table S2 Number and relative frequency of the destabilizing positively selected substitutions in the OPN4m and the OPN4x paralogs. 30 physicochemical properties were analysed in two categories, based on their nature: chemical and structural.
(PDF)

Table S3 Branch and branch-site selection tests and the respective estimated parameters. The asterisk (*) means that the alternative hypothesis is statistically significant at a 5% level, implementing the LRT (likelihood ratio test). Notes: *df* – degrees of freedom.
(PDF)

Table S4 Nucleotide substitution models and the respective estimated parameters for OPN4m, OPN4x and OPN4 alignments. Parameters: base frequencies, substitution ratio between the nucleotide bases (r), gamma shape parameter and proportion of invariable sites (p-inv). The comparison between the saturation index (I_{SS}) and the critical index value ($I_{SS.C}$) implemented by Xia et al. 2003 [80] were also represented, as well as the respective category of data saturation.
(PDF)

Acknowledgments

We thank João P. Machado and Kartik Sunagar for the helpful discussion and comments. Comments made by the Associate Editor and two anonymous reviewers improved an earlier version of the manuscript.

Author Contributions

Conceived and designed the experiments: AA RB. Performed the experiments: RB. Analyzed the data: RB AA. Contributed reagents/materials/analysis tools: AA WEJ SJO VV. Wrote the paper: RB WEJ AA.

References

1. Arendt D (2008) The evolution of cell types in animals: emerging principles from molecular studies. Nature reviews Genetics 9: 868–882.
2. Provencio I, Jiang G, De Grip WJ, Hayes WP, Rollag MD (1998) Melanopsin: An opsin in melanophores, brain, and eye. Proceedings of the National Academy of Sciences of the United States of America 95: 340–345.
3. Lucas RJ, Freedman MS, Munoz M, Garcia-Fernandez JM, Foster RG (2011) Regulation of the Mammalian Pineal by Non-rod, Non-cone, Ocular Photoreceptors. Science 505. doi:10.1126/science.284.5413.505.
4. Lucas RJ, Douglas RH, Foster RG (2001) Characterization of an ocular photopigment capable of driving pupillary constriction in mice. Nature neuroscience 4: 621–626.
5. Berson D, Dunn F, Takao M (2002) Phototransduction by retinal ganglion cells that set the circadian clock. Science (New York, NY) 295: 1070–1073.
6. Hattar S, Lucas RJ, Mrosovsky N, Thompson S, Douglas RH, et al. (2003) Melanopsin and rod-cone photoreceptive systems account for all major accessory visual functions in mice. Nature 424: 76–81.
7. Hankins MW, Lucas RJ (2002) The primary visual pathway in humans is regulated according to long-term light exposure through the action of a nonclassical photopigment. Current biology: CB 12: 191–198.
8. Bellingham J, Chaurasia SS, Melyan Z, Liu C, Cameron M a, et al. (2006) Evolution of melanopsin photoreceptors: discovery and characterization of a new melanopsin in nonmammalian vertebrates. PLoS biology 4: e254.
9. Pires SS, Shand J, Bellingham J, Arrese C, Turton M, et al. (2007) Isolation and characterization of melanopsin (Opn4) from the Australian marsupial Sminthopsis crassicaudata (fat-tailed dunnart). Proceedings Biological sciences/The Royal Society 274: 2791–2799.
10. Peirson S, Bovee-Geurts P, Lupi D, Jeffery G, DeGrip W, et al. (2004) Expression of the candidate circadian photopigment melanopsin (Opn4) in the mouse retinal pigment epithelium. Brain research Molecular brain research 123: 132–135.
11. Frigato E, Vallone D, Bertolucci C, Foulkes NS (2006) Isolation and characterization of melanopsin and pinopsin expression within photoreceptive sites of reptiles. Die Naturwissenschaften 93: 379–385.
12. Chaurasia SS, Rollag MD, Jiang G, Hayes WP, Haque R, et al. (2005) Molecular cloning, localization and circadian expression of chicken melanopsin (Opn4): differential regulation of expression in pineal and retinal cell types. Journal of neurochemistry 92: 158–170.
13. Davies WIL, Zheng L, Hughes S, Tamai TK, Turton M, et al. (2011) Functional diversity of melanopsins and their global expression in the teleost retina. Cellular and molecular life sciences: CMLS.
14. Terakita A (2005) The opsins. Genome biology: 1–9.
15. Davies WL, Foster RG, Hankins MW (2010) Focus on molecules: Melanopsin. Experimental eye research: 110–112.
16. Davies WL, Hankins MW, Foster RG (2010) Vertebrate ancient opsin and melanopsin: divergent irradiance detectors. Photochemical & photobiological sciences: Official journal of the European Photochemistry Association and the European Society for Photobiology 9: 1444–1457.
17. Koyanagi M, Kubokawa K, Tsukamoto H, Shichida Y, Terakita A (2005) Cephalochordate melanopsin: evolutionary linkage between invertebrate visual cells and vertebrate photosensitive retinal ganglion cells. Current biology: CB 15: 1065–1069.
18. Panda S, Nayak SK, Campo B, Walker JR, Hogenesch JB, et al. (2005) Illumination of the melanopsin signaling pathway. Science (New York, NY) 307: 600–604.
19. Peirson S, Foster RG (2006) Melanopsin: another way of signaling light. Neuron 49: 331–339.
20. Angueyra JM, Pulido C, Malagón G, Nasi E, Gomez MDP (2012) Melanopsin-expressing amphioxus photoreceptors transduce light via a phospholipase C signaling cascade. PloS one 7: e29813. 21.
21. Qiu X, Kumbalasiri T, Carlson SM, Wong KY, Krishna V, et al. (2005) Induction of photosensitivity by heterologous expression of melanopsin. Nature 433: 745–749.
22. Graham DM, Wong KY, Shapiro P, Frederick C, Pattabiraman K, et al. (2008) Melanopsin ganglion cells use a membrane-associated rhabdomeric phototransduction cascade. Journal of neurophysiology 99: 2522–2532.
23. Isoldi MC, Rollag MD, Castrucci AM de L, Provencio I (2005) Rhabdomeric phototransduction initiated by the vertebrate photopigment melanopsin. Proceedings of the National Academy of Sciences of the United States of America 102: 1217–1221.
24. Arendt D (2003) Evolution of eyes and photoreceptor cell types. The International journal of developmental biology 47: 563–571.
25. Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, et al. (2008) The amphioxus genome and the evolution of the chordate karyotype. Nature 453: 1064–1071.
26. Vandepoele K, De Vos W, Taylor JS, Meyer A, Van de Peer Y (2004) Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proceedings of the National Academy of Sciences of the United States of America 101: 1638–1643.
27. Rennison DJ, Owens GL, Taylor JS (2011) Opsin gene duplication and divergence in ray-finned fish. Molecular phylogenetics and evolution 62: 986–1008.
28. Gojobori J, Innan H (2009) Potential of fish opsin gene duplications to evolve new adaptive functions. Trends in genetics: TIG 25: 198–202.
29. Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Molecular biology and evolution 15: 568–573.
30. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution 24: 1586–1591.
31. Dong C, Zhang J, Qiao J, He G (2011) Positive Selection and Functional Divergence After Melanopsin Gene Duplication. Biochemical genetics.
32. Swanson WJ, Nielsen R, Yang Q (2003) Pervasive Adaptive Evolution in Mammalian Fertilization Proteins. Molecular Biology and Evolution 20: 18–20.
33. Wilcoxon F (1945) Individual comparisons by ranking methods. Biometrics Bulletin 1: 80–83.

34. König B, Arendt A, McDowell JH, Kahlert M, Hargrave P, et al. (1989) Three cytoplasmic loops of rhodopsin interact with transducin. Proceedings of the National Academy of Sciences of the United States of America 86: 6878–6882.
35. Woolley S, Johnson J, Smith M, Crandall K, McClellan D (2003) TreeSAAP: Selection on Amino Acid Properties using phylogenetic trees. Bioinformatics 19: 671–672.
36. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, et al. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic acids research 33: W299–302.
37. Gu X, Vander Velden K (2002) DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. Bioinformatics (Oxford, England) 18: 500–501.
38. Gu X (2001) Maximum-likelihood approach for gene family evolution under functional divergence. Molecular biology and evolution 18: 453–464.
39. Sgourakis NG, Bagos PG, Papasaikas PK, Hamodrakas SJ (2005) A method for the prediction of GPCRs coupling specificity to G-proteins using refined profile Hidden Markov Models. BMC bioinformatics 6: 104.
40. Arendt D, Wittbrodt J (2001) Reconstructing the eyes of Urbilateria. Philosophical transactions of the Royal Society of London Series B, Biological sciences 356: 1545–1563.
41. Melyan Z, Tarttelin EE, Bellingham J, Lucas RJ, Hankins MW (2005) Addition of human melanopsin renders mammalian cells photoresponsive. Nature 433: 741–745.
42. Spagnuolo a (2003) Unusual number and genomic organization of Hox genes in the tunicate Ciona intestinalis. Gene 309: 71–79.
43. Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, et al. (2002) The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. Science (New York, NY) 298: 2157–2167.
44. Kusakabe T, Kusakabe R, Kawakami I, Satou Y, Satoh N, et al. (2001) Ci-opsin1, a vertebrate-type opsin gene, expressed in the larval ocellus of the ascidian Ciona intestinalis. FEBS letters 506: 69–72.
45. Kusakabe T, Tsuda M (2003) Photoreceptive systems in ascidians. Photochemistry and photobiology 83: 248–252.
46. Holland PW, Garcia-Fernàndez J, Williams N a, Sidow a (1994) Gene duplications and the origins of vertebrate development. Development (Cambridge, England) Supplement 1994: 125–133.
47. Larhammar D, Nordström K, Larsson T a (2009) Evolution of vertebrate rod and cone phototransduction genes. Philosophical transactions of the Royal Society of London Series B, Biological sciences 364: 2867–2880.
48. Zhang J (2003) Evolution by gene duplication: an update. Trends in Ecology & Evolution 18: 292–298.
49. Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nature reviews Genetics 11: 97–108.
50. Vopalensky P, Kozmik Z (2009) Eye evolution: common use and independent recruitment of genetic components. Philosophical transactions of the Royal Society of London Series B, Biological sciences 364: 2819–2832.
51. Lamb TD, Pugh EN, Collin SP (2008) The Origin of the Vertebrate Eye. Evolution: Education and Outreach 1: 415–426.
52. Lamb TD, Collin SP, Pugh EN (2007) Evolution of the vertebrate eye: opsins, photoreceptors, retina and eye cup. Nature reviews Neuroscience 8: 960–976.
53. Jacobs GH (2009) Evolution of colour vision in mammals. Philosophical transactions of the Royal Society of London Series B, Biological sciences 364: 2957–2967.
54. Schmidt TM, Chen S-K, Hattar S (2011) Intrinsically photosensitive retinal ganglion cells: many subtypes, diverse functions. Trends in neurosciences 34: 572–580.
55. Peirson S, Halford S, Foster R (2009) The evolution of irradiance detection: melanopsin and the non-visual opsins. Philosophical transactions of the Royal Society of London Series B, Biological sciences 364: 2849–2865.
56. Heesy CP, Hall MI (2010) The nocturnal bottleneck and the evolution of mammalian vision. Brain, behavior and evolution 75: 195–203.
57. Bowmaker JK, Hunt DM (2006) Evolution of vertebrate visual pigments. Current biology: CB 16: R484–9.
58. Torii M, Kojima D, Okano T, Nakamura A, Terakita A, et al. (2007) Two isoforms of chicken melanopsins show blue light sensitivity. FEBS letters 581: 5327–5331.
59. Collin SP, Davies WL, Hart NS, Hunt DM (2009) The evolution of early vertebrate photoreceptors. Philosophical transactions of the Royal Society of London Series B, Biological sciences 364: 2925–2940.
60. Tsukamoto H, Terakita A (2010) Diversity and functional properties of bistable pigments. Photochemical & photobiological sciences: Official journal of the European Photochemistry Association and the European Society for Photobiology 9: 1435–1443.
61. Fritze O, Filipek S, Kuksa V, Palczewski K, Hofmann KP, et al. (2003) Role of the conserved NPxxY(x)5,6F motif in the rhodopsin ground state and during activation. Proceedings of the National Academy of Sciences of the United States of America 100: 2290–2295.
62. National Center for Biotechnology Information (NCBI). Available: http://www.ncbi.nlm.nih.gov/. Accessed: 26 September 2011.
63. Ensembl. Available: http://www.ensembl.org/index.html. Accessed: 26 September 2011.
64. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Matrix 32: 1792–1797. doi:10.1093/nar/gkh340.
65. Talavera G, Castresaba J (2007) Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. Systematic Biology 56: 564–577. doi:10.1080/10635150701472164.
66. Xia X, Xie Z, Salemi M, Chen L, Wang Y (2003) An index of substitution saturation and its application. Molecular phylogenetics and evolution 26: 1–7.
67. Posada D (2008) jModelTest: phylogenetic model averaging. Molecular biology and evolution 25: 1253–1256.
68. Guindon S, Gascuel O (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Systematic Biology 52: 696–704.
69. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics (Oxford, England) 17: 754–755.
70. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.
71. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783–791.
72. Muffato M, Louis A, Poisnel C-E, Roest Crollius H (2010) Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. Bioinformatics (Oxford, England) 26: 1119–1121.
73. Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Molecular biology and evolution 18: 1585–1592.
74. Doron-Faigenboim A, Stern A, Mayrose I, Bacharach E, Pupko T (2005) Selecton: a server for detecting evolutionary forces at a single amino-acid site. Bioinformatics (Oxford, England) 21: 2101–2103.
75. Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Molecular biology and evolution 19: 908–917.
76. Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Molecular biology and evolution 22: 2472–2479.
77. Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. Molecular biology and evolution 22: 1107–1118.
78. McClellan D a, Palfreyman EJ, Smith MJ, Moss JL, Christensen RG, et al. (2005) Physicochemical evolution and molecular adaptation of the cetacean and artiodactyl cytochrome b proteins. Molecular biology and evolution 22: 437–455.
79. Fiser A, Sali A (2003) Modeller: generation and refinement of homology-based protein structure models. Methods in enzymology 374: 461–491.
80. Shimamura T, Hiraki K, Takahashi N, Hori T, Ago H, et al. (2008) Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region. The Journal of biological chemistry 283: 17753–17756.
81. Murakami M, Kouyama T (2008) Crystal structure of squid rhodopsin. Nature 453: 363–367.
82. Schrödinger L. PyMOL Molecular Graphics System. Available: http://www.pymol.org/. Accessed: 6 July 2011.
83. Gu X (1999) Statistical methods for testing functional divergence after gene duplication. Molecular biology and evolution 16: 1664–1674.
84. Gu X (2006) A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences. Molecular biology and evolution 23: 1937–1945.

5. Final remarks

The molecular adaptation phenomenon is extremely important, enabling organisms to constantly adapt to different environmental pressures over time. Nevertheless, this process is extremely complex to access, since it is intended to know which modifications at the molecular level have led to the phenotypic changes in the organism and which enabled them to adapt to the existing selective pressures. Generally, the concept of molecular adaptation is understood as the integration of polymorphisms in a particular gene or genomic region, with advantageous functional implications to the organism [81–83]. However, the new genomic analyses as well as the study of gene families, have shown that the gain and loss of genes can also be an important factor in determining the observed phenotypic variation, that in most cases has a selective meaning [64, 79, 108, 109].

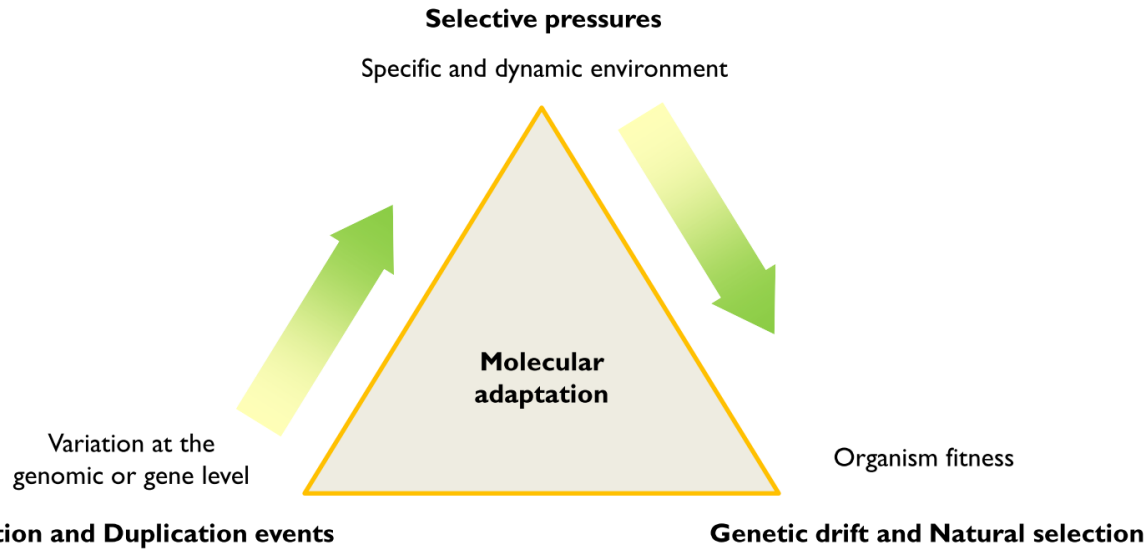


Figure 5.1. The concept of the adaptive molecular evolution. Variation at the genomic and gene level can be obtained by mutation or duplication events. Organism fitness depends on the environmental pressures in that specific period. The selective pressures are rather determinant in safeguarding or eliminating genetic variation either by genetic drift or natural selection.

Thus the study of the concept of molecular adaptation involves the integrated analysis of three important aspects (**figure 5.1**): (1) the source of variability, either by introduction of polymorphisms or by gain and loss of genes; (2) the change in environmental pressures determining the nature of the polymorphism (neutral, advantageous or disadvantageous) and

the destination of the duplicated genes; and (3) the action of natural selection and genetic drift deciding whether the polymorphism or the duplicated gene is introduced or lost [110, 111].

The sensorial receptors are excellent study objects to understand molecular adaptation: in one hand, they are the first intermediaries between the environmental stimulus and the organism, and on the other hand, they are potentially quite adaptable when the selective pressure changes. Therefore, the study of the sensorial gene families may enlighten the process of adaptation at the molecular level [26, 112]. Here, a photoreceptive sensorial gene was studied – the melanopsin (*OPN4*). The aim was to determine the molecular and evolutionary mechanisms that led to the adaptation of the melanopsin to different photic environments over time. Thus, taking into account our results, two evolutionary phenomena can be stated as particularly important to mediate the adaptive evolution of photoreceptive molecules:

5.1. Purifying or negative selection

It was been shown that melanopsins evolve generally under strong purifying or negative selection. This allowed us to conclude that melanopsins are extremely important from a functional standpoint, and the total or partial loss of its function will compromise the organism fitness [113]. Indeed, melanopsins are responsible for the regulation of circadian rhythms and it is therefore indirectly related to the basic processes of metabolism, body temperature regulation and the feeding, reproduction and hibernation processes [114]. On the other hand, it is easily understood why structural constraints are imposed to the acquisition of amino acid variability by melanopsins. As shown, mutations with structural implications are much less common than the chemical ones. This seems plausible since in this type of receptor, the ligand is unique and exclusive, the retinal (11-*cis*-retinal) [17, 18]. Thus, mutations with structural implications can compromise the pocket where the retinal is accommodated and therefore are negatively selected. Compared with other molecules, such as the chemoreceptors, this pattern should not be expected given the existence of an enormous amount of potential ligands in the environment [115, 116].

The heptatransmembrane structure and retinal-dependency are homologous characteristic of all opsins and they were maintained in this group only because certain domains remained unchanged, whereas others, still changing, have been properly compensated [117]. Therefore, taking into account the structural, physiological and sequence homologies within the opsin group, as well as their ancient emergence in the animal kingdom, one would expect the overall action of the negative selection to mediate the evolution of opsins.

5.2. Duplication events followed by minor episodes of positive selection

The phenomena of gene duplication are important generating new gene families. As we showed, melanopsins are represented by two paralogs (*OPN4m* and *OPN4x*) that arose through the 2R duplication event in the emergence of vertebrates. These events were extremely important as they contributed to the variety of visual pigments that currently can be found in the general plan of the vertebrate eye. In fact, the rhodopsins and conopsin gene representatives (*Rh1*, *Rh2*, *OPN1lw* and *OPN1sw*) [30] were likely to be originated during the 2R duplication event, as we have suggested for the melanopsin gene family. Moreover, the presence of genomic paralogons among vision-related genes produced by the 2R episode seems to be common pattern in visual opsins, as has been demonstrated through the study of the protein intervenes in the vertebrate visual cascade [76]. Obviously when the duplication event is mentioned, we are not referring to the exact moment of the duplication event, but to overall process that includes the selection, maintenance and accommodation of the copy (or copies) in the genome.

The phenomena of gain and loss of gene representatives of a certain gene family is also a form of adaptation, and gene content can be regulated simply by genetic drift or directional selection [71]. We showed that teleost fish genomes contain often five melanopsin copies in result of the 3R duplication event. The maintenance of such great number or copies may be related to the colonization of new photic environments by teleost fish during Carboniferous, where new copies could signal new quantities and/or qualities of light, specializing accordingly and improving the organism fitness [118]. Moreover, the *OPN4x* variant was loss in the mammalian lineage. Mammals are likely to have a nocturnal ancestry, and indeed, they are quite simple from the visual point of view: (1) mammals only contain the eye as

photoreception organ, (2) their eye show anatomical features associated with nocturnal or low light environments, and (3) they have a less number of visual and non-visual opsins [16]. The absence of the *OPN4x* paralog in the mammalian genomes might be related to the loss of a certain quality of light. Hypothetically, this environmental change could lead to the loss of photoreceptive genes by genetic drift, simply by integrating neutral non-sense mutations and subsequent miss-identification of the gene.

Maintaining gene copies in the genome could have an intermediate fitness-decreasing effect to the organism due to metabolic constraints, and hence some degree of specialization will have to occur in at least one of the copies, after duplication [74, 75]. Thus, it is expected, and our results have shown it, small episodes of positive selection or relaxed selective pressures after the duplication events. Such signatures can now be analyzed by looking at the nucleotide and amino acid patterns. Our results suggest type I and II functional divergence as well as positive selection between the *OPN4m* and *OPN4x*.

Firstly, (1) we found evidence of positive selection and functional divergence in the sites near the retinal pocket that are likely related to the spectral sensitivity of melanopsins [119] (**figure 5.2A**). This suggests that melanopsin *m* and *x* variants are sensible to different quantities or qualities of light. Furthermore, (2) we also found evidence of destabilizing positive selection in the second and third intracellular loops (**figure 5.2B**). These loops are involved in establishing the activation of a specific G-protein type, and are therefore responsible for the specificity of the light signalling process [120]. Since we found notorious amino acid variability in these loops, this may suggest some ambiguity in the G protein that interacts with melanopsins. Our predictions show that the Gio and Gq11 types are the most likely. Such results have functional significance because it suggests that melanopsins can regulate many intracellular processes and, consequently, be involved in more than one signal transduction process. This promiscuous behavior can be advantageous since melanopsins may carry out a greater number and/or more complex responses to light [121, 122]. This pattern was first described by us and it has not been reported so far, even in other opsins. Initial analysis carried out by us, have shown that rhodopsin is fairly conserved in these regions (data not shown), which does not suggest the ambiguous interaction with more than one G-protein, as we advance for melanopsins.

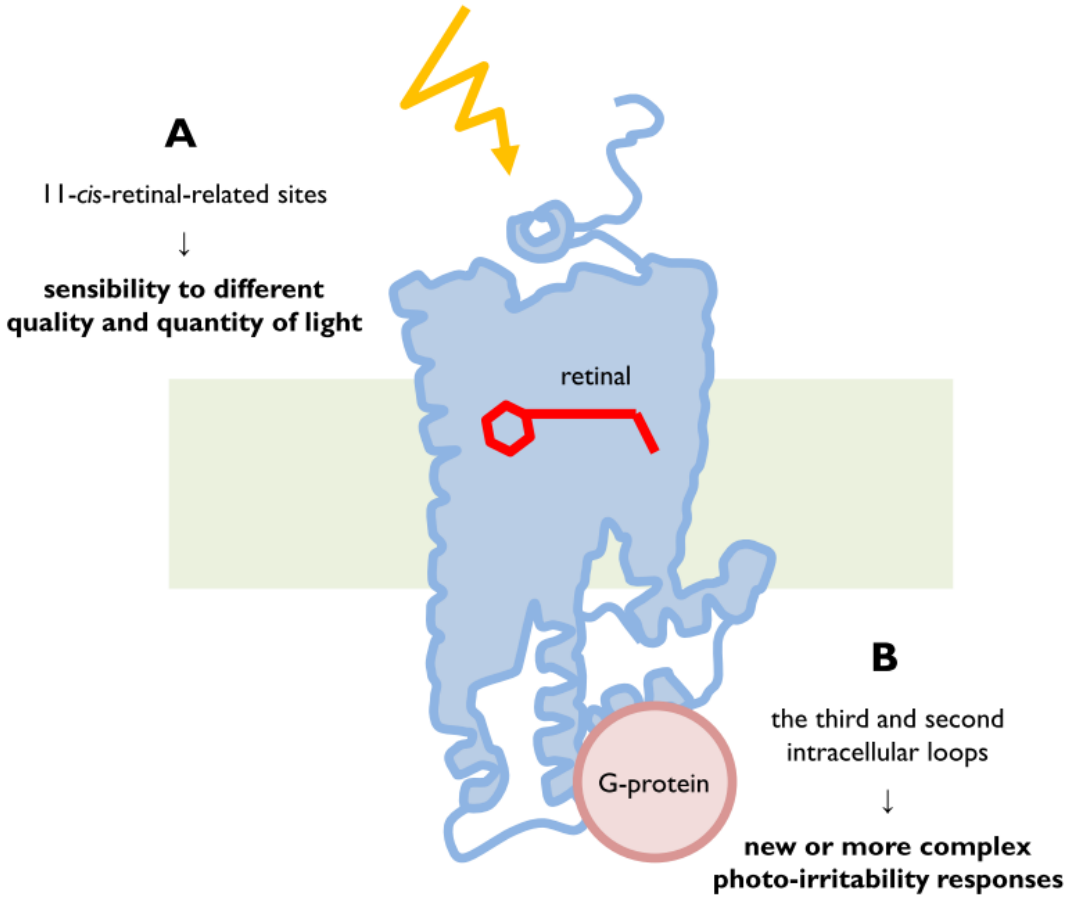


Figure 5.2. The adaptive domains in melanopsin. A. The retinal photo-isomerization-related residues are likely to be involved in melanopsins spectral sensitivity. B. The third and second intracellular loops establish the activation of a specific G-protein type, and are therefore responsible for the specificity of the light signalling process.

Thus, we suggested that melanopsins have diverged in function, on the one hand by specializing in different ranges of the light spectrum and, on the other hand establishing ambiguous intracellular interaction with different G-protein types (**figure 5.2**). The first mechanism has been previously mentioned in the literature as an engine of visual specialization in fish and birds [27, 123], but the second one was firstly reported by us.

5.3. Future perspectives

In the future, it would be interesting to extend this study to all opsins described in the vertebrate eyes, especially non-visual opsins, where there are no such studies. It would be important to realize how and when the gene gain and loss events happened, not only in the

opsin family during vertebrate evolution, but also in some specific groups as mammals and teleost fish, primates and bats, nocturnal and diurnal birds. The vertebrate group is very complex from a visual point of view: all non-mammalian vertebrates have other photoreception organs apart from the eye, birds are almost visual animals, using this sense in reproduction, with notorious implications in sexual selection; anthropoid primates have trichromatic color vision while all other mammals have dichromatic vision [10, 124]. Indeed, it would be important to compare the number and the pseudogenization events between these groups, in order to know how significant the gene content regulation was during the speciation process. Teleost fish group should also be further studied from the visual standpoint, particularly, to assess the number of photoreceptor genes they have in their genomes. As seen, teleost fish present five melanopsin copies. However, it is known that in higher vertebrates seventeen opsin genes were described, and thus in teleost fish this value should be much higher due to 3R duplication event [118].

On the other hand, it would be also important to verify if the adaptive mechanisms advanced by us (intra-family gene duplication, amino acid variation at the second and third intracellular loops and retinal-related sites) are the same in other opsin families. This analysis would also enable to assess if the same domains of the molecule are under different selective constraints in the different opsin families. Additionally, it would be interesting to study all the molecules involved in signal transduction pathway of light stimulus, implementing the approach used in this work. Thus, it would be possible to determine which proteins have more adaptive capacity: those that are found at the beginning of the cascade, or those that comprise the signal, which are in the final of the phototransduction pathway. This approach would provide an overview of the evolutionary mechanism that is responsible for the integration of the light stimuli, as well as to determine the involvement of protein-protein interactions in the signalization of light. This could be done in rhodopsins, for which the mechanism of signal transduction was previously characterized [125].

Finally, we concluded that the molecular evolution approach can be extremely useful in the inference of the mechanisms that allow molecules to adapt to different environmental conditions. Moreover, this approach also enables us to infer about conserved regions that assure the identity of protein families, establishing similar functions. We suggested that

melanopsins have diverged firstly, by regulating the number of genes in the different groups of vertebrates, and secondly, by signalizing different qualities of light and/or establishing ambiguous intracellular interactions with different G-protein types. Additionally, we showed that the functional and structural identity between melanopsins was maintained by strong purifying selection in certain domains of the molecule. These adaptive mechanisms, advanced by us, highlight not only the understanding of how the signalling light pathways implement the regulation of circadian rhythms, but also the comprehension of the adaptation to the changing photic environments.

References

1. Darwin C: *On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life.* London ;: Henry Frowde,; 1907.
2. Fernald RD: **Casting a genetic light on the evolution of eyes.** *Science (New York, N.Y.)* 2006, **313**:1914–8.
3. Hughes AL: **The origin of adaptive phenotypes**. *Proceedings of the National Academy of Sciences* 2008, **105**:13193–13194.
4. Yokoyama S, Tada T, Zhang H, Britt L: **Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates.** *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**:13480–5.
5. Cernuda-Cernuda R, DeGrip WJ, Cooper HM, Nevo E, García-Fernández JM: **The retina of Spalax ehrenbergi: novel histologic features supportive of a modified photosensory role.** *Investigative ophthalmology & visual science* 2002, **43**:2374–83.
6. David-Gray ZK, Cooper HM, Janssen JW, Nevo E, Foster RG: **Spectral tuning of a circadian photopigment in a subterranean "blind" mammal (Spalax ehrenbergi).** *FEBS letters* 1999, **461**:343–7.
7. Carvalho LS, Cowing JA, Wilkie SE, Bowmaker JK, Hunt DM: **The molecular evolution of avian ultraviolet- and violet-sensitive visual pigments.** *Molecular biology and evolution* 2007, **24**:1843–52.
8. Andersson S, Ornborg J, Andersson M: **Ultraviolet sexual dimorphism and assortative mating in blue tits**. *Proceedings of the Royal Society B: Biological Sciences* 1998, **265**:445–450.
9. Bright A, Waas JR: **Effects of bill pigmentation and UV reflectance during territory establishment in blackbirds**. *Animal Behaviour* 2002, **64**:207–213.
10. Collin SP, Davies WL, Hart NS, Hunt DM: **The evolution of early vertebrate photoreceptors.** *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 2009, **364**:2925–40.
11. Owens GL, Rennison DJ, Allison WT, Taylor JS: **In the four-eyed fish (Anableps anableps), the regions of the retina exposed to aquatic and aerial light do not express the same set of opsin genes.** *Biology letters* 2012, **8**:86–9.
12. Owens GL, Windsor DJ, Mui J, Taylor JS: **A fish eye out of water: ten visual opsins in the four-eyed fish, Anableps anableps.** *PloS one* 2009, **4**:e5970.
13. Lamb T, Collin S: **Evolution of the vertebrate eye: opsins, photoreceptors, retina and eye cup**. *Nature Reviews Neuroscience* 2007, **8**:960–976.
14. Yokoyama S, Yokoyama R: **Adaptive evolution of photoreceptors and visual pigments in vertebrates**. *Annual Review of Ecology and Systematics* 1996, **27**:543–567.
15. Peirson S, Halford S, Foster R: **The evolution of irradiance detection: melanopsin and the non-visual opsins.** *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 2009, **364**:2849–65.
16. Heesy CP, Hall MI: **The nocturnal bottleneck and the evolution of mammalian vision.** *Brain, behavior and evolution* 2010, **75**:195–203.
17. Terakita A: **The opsins**. *Genome biology* 2005:1–9.

18. Nickle B, Robinson PR: **The opsins of the vertebrate retina: insights from structural, biochemical, and evolutionary studies.** *Cellular and molecular life sciences : CMLS* 2007, **64**:2917–32.

19. Lohse M, Benovic J, Codina J, Caron M, Lefkowitz R: **beta-Arrestin: a protein that regulates beta-adrenergic receptor function**. *Science* 1990, **248**:1547–1550.

20. Attwood TK, Findlay JBC: **Fingerprinting G-protein-coupled receptors**. *"Protein Engineering, Design and Selection"* 1994, **7**:195–203.

21. Kolakowski LF: **GCRDb: a G-protein-coupled receptor database.** *Receptors & channels* 1994, **2**:1–7.

22. Yau K-W, Hardie RC: **Phototransduction motifs and variations.** *Cell* 2009, **139**:246–64.

23. Arendt D: **Evolution of eyes and photoreceptor cell types.** *The International journal of developmental biology* 2003, **47**:563–71.

24. Hisatomi O, Tokunaga F: **Molecular evolution of proteins involved in vertebrate phototransduction**. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 2002, **133**:509–522.

25. Bellingham J, Foster RG: **Opsins and mammalian photoentrainment.** *Cell and tissue research* 2002, **309**:57–71.

26. Yokoyama S, Tada T, Zhang H, Britt L: **Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates.** *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**:13480–5.

27. Carvalho LS, Cowing J a, Wilkie SE, Bowmaker JK, Hunt DM: **The molecular evolution of avian ultraviolet- and violet-sensitive visual pigments.** *Molecular biology and evolution* 2007, **24**:1843–52.

28. Fernald RD: **Casting a genetic light on the evolution of eyes.** *Science (New York, N.Y.)* 2006, **313**:1914–8.

29. Lamb TD: **Evolution of vertebrate retinal photoreception.** *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 2009, **364**:2911–24.

30. Shichida Y, Matsuyama T: **Evolution of opsins and phototransduction.** *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 2009, **364**:2881–95.

31. Berson DM: **Strange vision: ganglion cells as circadian photoreceptors.** *Trends in neurosciences* 2003, **26**:314–20.

32. Hattar S, Liao HW, Takao M, Berson DM, Yau KW: **Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity.** *Science (New York, N.Y.)* 2002, **295**:1065–70.

33. Hatori M, Panda S: **The emerging roles of melanopsin in behavioral adaptation to light.** *Trends in molecular medicine* 2010, **16**:435–46.

34. Bailes HJ, Lucas RJ: **Melanopsin and inner retinal photoreception.** *Cellular and molecular life sciences : CMLS* 2010, **67**:99–111.

35. Gomez M del P, Angueyra JM, Nasi E: **Light-transduction in melanopsin-expressing photoreceptors of Amphioxus.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**:9081–6.

36. Torii M, Kojima D, Okano T, Nakamura A, Terakita A, Shichida Y, Wada A, Fukada Y: **Two isoforms of chicken melanopsins show blue light sensitivity.** *FEBS letters* 2007, **581**:5327–31.

37. Lucas RJ, Douglas RH, Foster RG: **Characterization of an ocular photopigment capable of driving pupillary constriction in mice.** *Nature neuroscience* 2001, **4**:621–6.

38. Provencio I, Jiang G, De Grip WJ, Hayes WP, Rollag MD: **Melanopsin: An opsin in melanophores, brain, and eye.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**:340–5.

39. Freedman MS: **Regulation of Mammalian Circadian Behavior by Non-rod, Non-cone, Ocular Photoreceptors**. *Science* 1999, **284**:502–504.

40. Provencio I, Rodriguez IR, Jiang G, Hayes WP, Moreira EF, Rollag MD: **A novel human opsin in the inner retina.** *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2000, **20**:600–5.

41. Semo M, Muñoz Llamosas M, Foster RG, Jeffery G: **Melanopsin (Opn4) positive cells in the cat retina are randomly distributed across the ganglion cell layer.** *Visual neuroscience* 2005, **22**:111–6.

42. Hannibal J, Hindersson P, Nevo E, Fahrenkrug J: **The circadian photopigment melanopsin is expressed in the blind subterranean mole rat, Spalax**. *Neuroreport* 2002, **13**:1411–1414.

43. Do MTH, Kang SH, Xue T, Zhong H, Liao H-W, Bergles DE, Yau K-W: **Photon capture and signalling by melanopsin retinal ganglion cells.** *Nature* 2009, **457**:281–7.

44. Murakami M, Kouyama T: **Crystal structure of squid rhodopsin.** *Nature* 2008, **453**:363–7.

45. Koyanagi M, Kubokawa K, Tsukamoto H, Shichida Y, Terakita A: **Cephalochordate melanopsin: evolutionary linkage between invertebrate visual cells and vertebrate photosensitive retinal ganglion cells.** *Current biology : CB* 2005, **15**:1065–9.

46. Hardie RC: **Phototransduction in Drosophila melanogaster**. 2001, **3409**:3403–3409.

47. Panda S, Nayak SK, Campo B, Walker JR, Hogenesch JB, Jegla T: **Illumination of the melanopsin signaling pathway.** *Science (New York, N.Y.)* 2005, **307**:600–4.

48. Qiu X, Kumbalasiri T, Carlson SM, Wong KY, Krishna V, Provencio I, Berson DM: **Induction of photosensitivity by heterologous expression of melanopsin.** *Nature* 2005, **433**:745–9.

49. Isoldi MC, Rollag MD, Castrucci AM de L, Provencio I: **Rhabdomeric phototransduction initiated by the vertebrate photopigment melanopsin.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**:1217–21.

50. Angueyra JM, Pulido C, Malagón G, Nasi E, Gomez MDP: **Melanopsin-expressing amphioxus photoreceptors transduce light via a phospholipase C signaling cascade.** *PloS one* 2012, **7**:e29813.

51. Davies WL, Foster RG, Hankins MW: **Focus on molecules: Melanopsin.** *Experimental eye research* 2010:110–112.

52. Davies WL, Hankins MW, Foster RG: **Vertebrate ancient opsin and melanopsin: divergent irradiance detectors.** *Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology* 2010, **9**:1444–57.

53. Koyanagi M, Terakita A: **Gq-coupled rhodopsin subfamily composed of invertebrate visual pigment and melanopsin.** *Photochemistry and photobiology* 2008, **84**:1024–30.

54. Bellingham J, Chaurasia SS, Melyan Z, Liu C, Cameron M a, Tarttelin EE, Iuvone PM, Hankins MW, Tosini G, Lucas RJ: **Evolution of melanopsin photoreceptors: discovery and characterization of a new melanopsin in nonmammalian vertebrates.** *PLoS biology* 2006, **4**:e254.

55. Davies WIL, Zheng L, Hughes S, Tamai TK, Turton M, Halford S, Foster RG, Whitmore D, Hankins MW: **Functional diversity of melanopsins and their global expression in the teleost retina.** *Cellular and molecular life sciences : CMLS* 2011.

56. Peirson S, Bovee-Geurts P, Lupi D, Jeffery G, DeGrip W, Foster R: **Expression of the candidate circadian photopigment melanopsin (Opn4) in the mouse retinal pigment epithelium.** *Brain research. Molecular brain research* 2004, **123**:132–5.

57. Frigato E, Vallone D, Bertolucci C, Foulkes NS: **Isolation and characterization of melanopsin and pinopsin expression within photoreceptive sites of reptiles.** *Die Naturwissenschaften* 2006, **93**:379–85.

58. Chaurasia SS, Rollag MD, Jiang G, Hayes WP, Haque R, Natesan a, Zatz M, Tosini G, Liu C, Korf HW, Iuvone PM, Provencio I: **Molecular cloning, localization and circadian expression of chicken melanopsin (Opn4): differential regulation of expression in pineal and retinal cell types.** *Journal of neurochemistry* 2005, **92**:158–70.

59. Arendt D: **The evolution of cell types in animals: emerging principles from molecular studies.** *Nature reviews. Genetics* 2008, **9**:868–82.

60. Arendt D, Wittbrodt J: **Reconstructing the eyes of Urbilateria.** *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 2001, **356**:1545–63.

61. **National Center for Biotechnology Information** [Http://www.ncbi.nlm.nih.gov/].

62. **ENSEMBL Genome Browser** [http://www.ensembl.org/index.html].

63. Holland PW, Garcia-Fernàndez J, Williams N a, Sidow a: **Gene duplications and the origins of vertebrate development.** *Development (Cambridge, England). Supplement* 1994, **1994**:125–33.

64. Vandepoele K, De Vos W, Taylor JS, Meyer A, Van de Peer Y: **Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**:1638–43.

65. Ohno S: *Evolution by gene duplication. S. Ohno. Springer-Verlag, New York. 1970. 160 pp*. Springer-Verlag; 1970, **9**:160.

66. Ohta T: **Further Simulation Studies on Evolution by Gene Duplication**. *Evolution* 1988, **42**:375.

67. Hughes AL: *Adaptive Evolution of Genes and Genomes*. Oxford University Press; 1999:288.

68. Zhang J: **Evolution by gene duplication: an update**. *Trends in Ecology & Evolution* 2003, **18**:292–298.

69. Lynch M: **The Evolutionary Fate and Consequences of Duplicate Genes**. *Science* 2000, **290**:1151–1155.

70. Prince VE, Pickett FB: **Splitting pairs: the diverging fates of duplicated genes.** *Nature reviews. Genetics* 2002, **3**:827–37.

71. Innan H, Kondrashov F: **The evolution of gene duplications: classifying and distinguishing between models.** *Nature reviews. Genetics* 2010, **11**:97–108.

72. Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J: **Preservation of Duplicate Genes by Complementary, Degenerative Mutations**. 1999.

73. Lynch M, Force a: **The probability of duplicate gene preservation by subfunctionalization.** *Genetics* 2000, **154**:459–73.

74. Lynch M, O'Hely M, Walsh B, Force a: **The probability of preservation of a newly arisen gene duplicate.** *Genetics* 2001, **159**:1789–804.

75. Papp B, Pál C, Hurst LD: **Dosage sensitivity and the evolution of gene families in yeast.** *Nature* 2003, **424**:194–7.

76. Larhammar D, Nordström K, Larsson T a: **Evolution of vertebrate rod and cone phototransduction genes.** *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 2009, **364**:2867–80.

77. Hokamp K, McLysaght A, Wolfe KH: **The 2R hypothesis and the human genome sequence.** *Journal of structural and functional genomics* 2003, **3**:95–110.

78. Holland PW, Garcia-Fernàndez J, Williams NA, Sidow A: **Gene duplications and the origins of vertebrate development.** *Development (Cambridge, England). Supplement* 1994:125–33.

79. Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu J-K, Benito-Gutiérrez EL, Dubchak I, Garcia-Fernàndez J, Gibson-Brown JJ, Grigoriev I V, Horton AC, De Jong PJ, Jurka J, Kapitonov V V, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio L a, Salamov A a, Satou Y, Sauka-Spengler T, Schmutz J, Shin-I T, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PWH, Satoh N, Rokhsar DS: **The amphioxus genome and the evolution of the chordate karyotype.** *Nature* 2008, **453**:1064–71.

80. Gojobori J, Innan H: **Potential of fish opsin gene duplications to evolve new adaptive functions.** *Trends in genetics : TIG* 2009, **25**:198–202.

81. Balaresque PL, Ballereau SJ, Jobling MA: **Challenges in human genetic diversity: demographic history and adaptation.** *Human molecular genetics* 2007, **16 Spec No**:R134–9.

82. Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, J Sninsky J, Adams MD, Cargill M: **A scan for positively selected genes in the genomes of humans and chimpanzees.** *PLoS biology* 2005, **3**:e170.

83. Swanson WJ: **Adaptive evolution of genes and gene families**. *Current Opinion in Genetics & Development* 2003, **13**:617–622.

84. Klein J, Sato A, Nagl S, O'hUigin C: **Molecular trans-species polymorphism**. *Annual Review of Ecology and Systematics* 1998, **29**:1–21.

85. Bamshad M, Wooding SP: **Signatures of natural selection in the human genome.** *Nature reviews. Genetics* 2003, **4**:99–111.

86. Nielsen R: **Molecular signatures of natural selection.** *Annual review of genetics* 2005, **39**:197–218.

87. Kimura M: **The Neutral Theory of Molecular Evolution**. *Scientific American* 1979, **241**:98–126.

88. Yang Z, Bielawski JP: **Statistical methods for detecting molecular adaptation**. *Trends in Ecology & Evolution* 2000, **15**:496–503.

89. Hughes AL, Nei M: **Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection.** *Nature* 1988, **335**:167–70.

90. Yang Z: **Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution.** *Molecular biology and evolution* 1998, **15**:568–73.

91. Hughes AL, Nei M: **Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection.** *Nature* 1988, **335**:167–70.

92. Yang Z, Nielsen R: **Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages.** *Molecular biology and evolution* 2002, **19**:908–17.

93. Doron-Faigenboim A, Stern A, Mayrose I, Bacharach E, Pupko T: **Selecton: a server for detecting evolutionary forces at a single amino-acid site.** *Bioinformatics (Oxford, England)* 2005, **21**:2101–3.

94. Nielsen R, Yang Z: **Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene.** *Genetics* 1998, **148**:929–36.

95. Yang Z, Nielsen R, Goldman N, Pedersen a M: **Codon-substitution models for heterogeneous selection pressure at amino acid sites.** *Genetics* 2000, **155**:431–49.

96. Zhang J, Nielsen R, Yang Z: **Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level.** *Molecular biology and evolution* 2005, **22**:2472–9.

97. McClellan DA, McCracken KG: **Estimating the Influence of Selection on the Variable Amino Acid Sites of the Cytochrome b Protein Functional Domains**. *Molecular Biology and Evolution* 2001, **18**:917–925.

98. McClellan D a, Palfreyman EJ, Smith MJ, Moss JL, Christensen RG, Sailsbery JK: **Physicochemical evolution and molecular adaptation of the cetacean and artiodactyl cytochrome b proteins.** *Molecular biology and evolution* 2005, **22**:437–55.

99. McClellan DA, Palfreyman EJ, Smith MJ, Moss JL, Christensen RG, Sailsbery JK: **Physicochemical evolution and molecular adaptation of the cetacean and artiodactyl cytochrome b proteins.** *Molecular biology and evolution* 2005, **22**:437–55.

100. Porter ML, Cronin TW, McClellan DA, Crandall KA: **Molecular characterization of crustacean visual pigments and the evolution of pancrustacean opsins.** *Molecular biology and evolution* 2007, **24**:253–68.

101. Osorio DS, Antunes A, Ramos MJ: **Structural and functional implications of positive selection at the primate angiogenin gene.** *BMC evolutionary biology* 2007, **7**:167.

102. Marques AT, Antunes A, Fernandes PA, Ramos MJ: **Comparative evolutionary genomics of the HADH2 gene encoding Abeta-binding alcohol dehydrogenase/17beta-hydroxysteroid dehydrogenase type 10 (ABAD/HSD10).** *BMC genomics* 2006, **7**:202.

103. Lichtarge O, Bourne HR, Cohen FE: **An evolutionary trace method defines binding surfaces common to protein families.** *Journal of molecular biology* 1996, **257**:342–58.

104. Gu X: **Statistical methods for testing functional divergence after gene duplication.** *Molecular biology and evolution* 1999, **16**:1664–74.

105. Gu X: **Maximum-likelihood approach for gene family evolution under functional divergence.** *Molecular biology and evolution* 2001, **18**:453–64.

106. Gu X: **A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences.** *Molecular biology and evolution* 2006, **23**:1937–45.

107. Golding GB, Dean AM: **The structural basis of molecular adaptation**. *Molecular Biology and Evolution* 1998, **15**:355–369.

108. Spagnuolo a: **Unusual number and genomic organization of Hox genes in the tunicate Ciona intestinalis**. *Gene* 2003, **309**:71–79.

109. Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, Harafuji N, Hastings KEM, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen I a, Necula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang H-G, Awazu S, Azumi K, Boore J, Branno M, Chin-Bow S, DeSantis R, Doyle S, Francino P, Keys DN, Haga S, Hayashi H, Hino K, Imai KS, Inaba K, Kano S, Kobayashi K, Kobayashi M, Lee B-I, Makabe KW, Manohar C, Matassi G, Medina M,

Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H, Ohta F, Oishi K, Rigoutsos I, Sano M, Sasaki A, Sasakura Y, Shoguchi E, Shin-i T, Spagnuolo A, Stainier D, Suzuki MM, Tassy O, Takatori N, Tokuoka M, Yagi K, Yoshizaki F, Wada S, Zhang C, Hyatt PD, Larimer F, Detter C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS: **The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins.** *Science (New York, N.Y.)* 2002, **298**:2157–67.

110. Yokoyama S: **Evaluating adaptive evolution.** *Nature genetics* 2002, **30**:350–1.
111. Kondrashov F a, Rogozin IB, Wolf YI, Koonin E V: **Selection in the evolution of gene duplications.** *Genome biology* 2002, **3**:RESEARCH0008.
112. Dong D, Jones G, Zhang S: **Dynamic evolution of bitter taste receptor genes in vertebrates.** *BMC evolutionary biology* 2009, **9**:12.
113. Hurst LD, Smith NGC: **Do essential genes evolve slowly?** *Current Biology* 1999, **9**:747–750.
114. Schmidt TM, Chen S-K, Hattar S: **Intrinsically photosensitive retinal ganglion cells: many subtypes, diverse functions.** *Trends in neurosciences* 2011, **34**:572–580.
115. Tokuriki N, Tawfik DS: **Stability effects of mutations and protein evolvability.** *Current opinion in structural biology* 2009, **19**:596–604.
116. Aravind L, Mazumder R, Vasudevan S, Koonin E V: **Trends in protein evolution inferred from sequence and structure analysis**. *Current Opinion in Structural Biology* 2002, **12**:392–399.
117. Camps M, Herman A, Loh E, Loeb LA: **Genetic constraints on protein evolution.** *Critical reviews in biochemistry and molecular biology* 2007, **42**:313–26.
118. Rennison DJ, Owens GL, Taylor JS: **Opsin gene duplication and divergence in ray-finned fish.** *Molecular phylogenetics and evolution* 2011, **62**:986–1008.
119. Bowmaker JK, Hunt DM: **Evolution of vertebrate visual pigments.** *Current biology : CB* 2006, **16**:R484–9.
120. König B, Arendt A, McDowell JH, Kahlert M, Hargrave P, Hofmann KP: **Three cytoplasmic loops of rhodopsin interact with transducin.** *Proceedings of the National Academy of Sciences of the United States of America* 1989, **86**:6878–82.
121. Khersonsky O, Roodveldt C, Tawfik DS: **Enzyme promiscuity: evolutionary and mechanistic aspects.** *Current opinion in chemical biology* 2006, **10**:498–508.
122. Kondrashov FA: **In search of the limits of evolution.** *Nature genetics* 2005, **37**:9–10.
123. Terai Y, Seehausen O, Sasaki T, Takahashi K, Mizoiri S, Sugawara T, Sato T, Watanabe M, Konijnendijk N, Mrosso HDJ, Tachida H, Imai H, Shichida Y, Okada N: **Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids.** *PLoS biology* 2006, **4**:e433.
124. Bowmaker JK: **Evolution of vertebrate visual pigments.** *Vision research* 2008, **48**:2022–41.
125. Hamer RD, Nicholas SC, Tranchina D, Lamb TD, Jarvinen JLP: **Toward a unified model of vertebrate rod phototransduction.** *Visual neuroscience* 2005, **22**:417–36.